Assessing spring phenology of a temperate woodland: a multiscale comparison of ground, Unmanned Aerial Vehicle and Landsat satellite observations

Abstract. The monitoring of forest phenology in a cost-effective manner, at a fine spatial scale and over relatively large areas remains a significant challenge. To address this issue, unmanned aerial vehicles (UAVs) appear to be a potential new platform for forest phenology monitoring. This article assesses the potential of UAV data to track the temporal dynamics of spring phenology, from the individual tree to woodland scale, and cross-compare UAV results against ground and satellite observations, in order to better understand characteristics of UAV data and assess potential for use in validation of satellite-derived phenology. A time series of UAV data (5 cm spatial resolution, ~7 day temporal resolution) were acquired in tandem with an intensive ground campaign during the spring season of 2015 across a 15 ha mixed woodland. Phenophase transition dates were estimated at an individual tree-level using UAV time series of Normalized Difference Vegetation Index (NDVI) and Green Chromatic Coordinate (GCC) and validated against visual observations of tree phenology. UAV-derived start of season dates could be predicted with an accuracy of less than 1 week. The analysis was scaled to a plot level, where ground (visual assessment and understorey development), UAV and Landsat metrics were compared, indicating UAV data is effective for tracking canopy phenology, as opposed to ecosystem dynamics detected by satellites. The UAV data were used to automatically map phenological events for individual trees across the whole woodland, demonstrating that contrasting canopy phenological events can occur within the extent of a single Landsat pixel. This, and a large temporal gap in the Landsat series, accounted for the poor relationships found between UAV- and Landsat-derived phenometrics (R^2<0.50) in this study. An opportunity is

now available to track very fine scale land surface changes over contiguous vegetation communities, providing information which could improve characterization of vegetation phenology at multiple scales.

Keywords: drone, consumer-grade camera, land surface phenology, forest phenology, individual tree level.

1. Introduction

Plant phenological events influence carbon, energy and water cycles within terrestrial ecosystems (Garrity et al. 2011; Mizunuma et al. 2013), operating from local to global scales, as around 55% of the Earth`s land surface is covered by grasslands, shrub lands and forests (Bartholomé and Belward 2005). As plant phenology events are highly sensitive to climate fluctuations, the timing of these events has been used as an independent indicator of climate change (Menzel et al. 2006; Thackeray et al. 2010), mainly in temperate environments (Fisher and Mustard 2007). A changing climate can drive shifts in plant phenology, with potential impacts on ecosystem services, ecosystem dynamics, plant-based economies, trophic interactions and species ranges (Campoy et al. 2011; Morisette et al. 2009; Sparks 2014; White et al. 2009). Assessing and monitoring phenological dynamics at various ecological scales are therefore key requirements to improve understanding of how plants respond to a changing world and how this influences forest ecosystems (Moore et al. 2016; Morisette et al. 2009).

Remote sensing techniques have been used to monitor vegetation phenology to complement traditional ground based manual measurements (Polgar and Primack 2013). Currently, vegetation phenology monitoring by remote sensing is effectively performed at two

contrasting scales: by ground and near-surface remote sensing or by coarse spatial resolution satellite sensors. The great advantage of satellite sensors is the capability to detect continuous patterns of vegetation changes across the land surface (Eastman et al. 2013; Rodriguez-Galiano et al. 2015). However, the estimation of the timing of the phenology events at a tree species level has many uncertainties, mainly due to the absence of validation data and the relatively coarse spatial resolution of many satellite remote sensing instruments, which often integrate the spectral response of many species (Eastman et al. 2013; Hmimina et al. 2013), each with a particular phenology (Melaas et al. 2013; Polgar and Primack 2011), therefore limiting the phenological representativeness at species-level (Delbart et al. 2005). Some of these scale-based limitations could be addressed by using medium spatial resolution satellites, but image availability can be significantly reduced due to cloud contamination, reducing the already lower dataset temporal resolution (compared to coarse spatial resolution) and increasing prediction uncertainties (Melaas et al. 2013; White et al. 2014). Combining imagery from multiple sensors such as Landsat and Sentinel-2 (Wang et al. 2017) could improve the quality of observations at this scale. However, even with an appropriate temporal resolution dataset, mixed vegetation composition is still an issue at a 30 m spatial resolution (Fisher et al. 2006; Liu et al. 2017).

On the other hand, ground (e.g. upward-pointing digital camera (Ryu et al. 2012)) and near-surface (e.g. tower-mounted camera (Brown et al. 2017)) sensors have the ability to observe individual plants at a daily or sub-daily frequency, detecting very subtle variations in vegetation phenology, which in turn makes it possible to accurately estimate phenological metrics (Ryu et al. 2012; Zhao et al. 2012), but over a limited area. Nevertheless, ecosystem representativeness can potentially be increased by phenological networks of ground and near-surface sensors (Nasahara and Nagai 2015) and the derived data set can allow more objective and direct comparisons with spaceborne measures (Baumann et al. 2017) than is possible with traditional ground based manual measurements. Despite these advantages, such networks still

present issues related to viewing angle, domination of the field of view by trees closest to the sensor and areal representativeness, which can confound the comparisons (Hmimina et al. 2013; Hufkens et al. 2012). While satellite sensors have synoptic views (Liu et al. 2017), near-surface sensors are often oriented with a view angle that is close to be horizontal (Liu et al. 2017), which may result in near-surface sensors receiving a smaller contribution from the background cover, such as snow and understorey vegetation, (Hufkens et al. 2012; Liu et al. 2017; Mizunuma et al. 2013) and higher contribution from the leaf layers under the canopy top (Keenan et al. 2014). Secondly, because ground/near-surface data are usually not georeferenced, an assumption must be made that such data are representative of the satellite pixel(s) area (Zhang et al. 2017).

An intermediate level of observation, between ground/near-surface and spaceborne data, can be achieved by airborne sensors (Higgins et al. 2011). However, an evident constraint is the use of manned aircraft, which implies high operational/logistical costs (Anderson and Gaston 2013; Hill et al. 2010), which make it difficult to perform frequent flights needed in phenology studies. The monitoring of vegetation phenology in a cost-effective manner, at a fine spatial scale and over relatively large areas therefore remains a significant challenge (Hufkens et al. 2012; Morris et al. 2013). To address this issue, unmanned aerial vehicles (UAVs) appear as a potential new option for vegetation phenology monitoring (Berra et al. 2016; Burkart et al. 2017; Dandois and Ellis 2013; Klosterman et al. 2018; Klosterman and Richardson 2017). UAVs offer scientists new opportunities for scale-appropriate measurement of ecological phenomena, delivering fine spatial resolution data at user-controlled revisit periods with relatively low cost (Anderson and Gaston 2013). Therefore, convenient temporal resolutions can be planned with UAVs in order to provide appropriate time-series to monitor phenological changes across diverse spatial scales (Berra et al. 2016; Dandois and Ellis 2013; Klosterman et al. 2018).

Advances in use of UAVs have been possible due not only to technological developments in UAVs (including positioning systems and sensors), but also to significant advances in data processing techniques, especially in (digital) photogrammetry and computer vision (Colomina and Molina 2014). Traditional photogrammetry (Tsingas 1992) proved not to be ideal for processing blocks of UAV images due to the irregularity of such images, which, in contrast, is no obstacle for approaches based on Structure-from-Motion (SfM) (Snavely et al. 2008). While the SfM approach has a number of advantages, it equally has a number of data collection (e.g. high overlap requirements) and data processing (e.g. artifacts) challenges; specifically, areas with dense vegetation are difficult targets for accurate feature matching and consequently topographic reconstruction (James et al. 2017; Woodget et al. 2017). Nevertheless, with the continuous interest in UAV-sourced images and SfM, the method is expected to continue to develop (James et al. 2017).

A few recent studies have shown that UAV time series data can detect the seasonal profile of deciduous forests in a manner similar to satellite sensors (Berra et al. 2017; Dandois and Ellis 2013), suggesting (without validating) that UAV time series can be useful to validate satellite-based phenological products. A couple of studies have explored the potential of UAV data to track individual tree-level phenology. We have previously conducted an exploratory study where a time series of a UAV colour index successfully tracked canopy seasonal changes of four individual oak trees (springtime), apparently matching visual observations of leaf development (Berra et al. 2016). Klosterman and Richardson (2017) monitored a complete seasonal cycle of 30 individual deciduous trees with UAV colour indices, where UAV derived phenometrics were successfully validated against ground observations of leaf development. More recently, Klosterman et al. (2018) used time series of UAV colour indices to track the seasonality of a mixed forest at a community level (10 x 10 m grid), whilst investigating how

this fine-scale perspective relates to land surface phenology (LSP) from coarser spatial resolution satellite sensors.

Despite these first insights, a number of challenges remain to be addressed in order to advance forest phenology monitoring with UAV data. Firstly, individual tree level analysis might be hampered by errors in the spatial alignment of tree crowns across acquisition dates (Klosterman et al. 2018), making a multi-scale investigation from the individual to landscape level difficult. Secondly, effectively decoupling understory and canopy phenology is challenging due to the high spectral variability within forested areas, especially from very high spatial resolution UAV data (Lin et al. 2018). Thirdly, the radiometric calibration of UAV imagery can be difficult, especially from consumer-grade cameras and from multi-temporal data sets experiencing severe changes in illumination conditions within and across acquisition dates (Berra et al. 2017). Fourthly, there is a need to move analysis beyond solely visible wavelengths. The last two points are particularly important to improve the ability to compare the UAV results with LSP derived from well calibrated and multi-band satellite data.

The aim of this research is to assess the potential of UAV data to track the temporal dynamics of spring phenology, from the individual tree to woodland scale, and to cross-compare UAV results against ground and satellite observations, in order to better understand characteristics of UAV data and assess potential for use in validation of satellite-derived phenology. The challenge of accurately monitoring the phenological behaviour of individual organisms over contiguous vegetation communities (Hufkens et al. 2012; Morris et al. 2013; Tang et al. 2016) is addressed, which can provide novel insights into species phenology and a better understanding of relationships between ecosystem processes (as observed with satellite sensors) and species-specific phenological events (as observed on the ground and by near-surface sensors). Specifically, we expand previous efforts of using UAV data for forest

phenology by: 1) validating UAV phenology against a larger number of ground observations (120 trees, 40 of which are evergreen), 2) testing the potential of NIR-based spectral VIs to detect phenology, in complement to colour indices, 3) testing methodological options to diminish the influence of understorey and/or shadows in the response of canopy phenology (detecting canopy rather than ecosystem phenology), 4) examining the spatial and temporal characteristics of tree species-specific phenology across an entire woodland (>4000 trees), and 5) assessing and better understanding the fine-scale spatial variability in phenology events occurring at a sub-pixel level for widely-used satellite data sets.

2. Methodology

2.1 Study area and forest inventory

The study area consists of around15 ha of mixed deciduous and conifer woodland surrounded by agricultural fields at Cockle Park Farm, located in the northeast of England (55.219867° N, 1.698661° W), around 37 km from Newcastle upon Tyne. The terrain within the woods has relatively flat topography with an altitude of approximately 75 m above sea level. A temperate climate is observed with a mean annual temperature of 8.3 °C (Newcastle University 2015). The site was chosen because: i) it offers diversity in terms of tree species composition ii) it has easy access; iii) it is large enough to be imaged by several Landsat pixels and, most importantly; iv) it allowed repeated UAV flights to be undertaken with minimal risk, as safety is paramount in UAV operations. Ground and UAV data were collected only during the spring season of 2015, as collecting data for a longer period of time was not feasible within the timeframe of this study (due to personnel availability).

Six plots were installed in order to sample individuals from the main tree species within this woodland (as shown in Figure 5a): European larch (*Larix decidua*), Sycamore (*Acer pseudoplatanus* L.), Sessile oak (*Quercus petrea*), Sitka spruce (*Picea sitchensis*), Norway spruce (*Picea abies*) and English oak (*Quercus robur*). A sample size of 20 trees per plot was adopted (Liang and Schwartz 2009), totalling 120 individuals. The ground within the Sitka and Norway spruce plots was mostly covered by litterfall, whilst a mix of herbs and grasses were predominant in the other plots. An overview of the plots along with the tree species occurring at each plot can be seen in the supplementary material (Figure S1).

Geographical coordinates of the selected trees were surveyed with total station and Global Navigation Satellite System (GNSS) stations in order to aid in the ground vs remote sensing data comparisons. Diameter at breast height (DBH) was measured with a tape measure at a height of 1.3 m. Tree height was measured using a vertex hypsometer (Vertex IV) and transponder (Transponder T3) (Haglof Sweden AB, Langsele, Sweden). Except for the Norway spruce plot, a large variability in DBH (from ±4.6 to ±12.0 cm) and height (from ±0.8 to ±4.5 m) occurred (Figure S1), and this is due to uneven-aged trees growing within the same plot.

2.2 Visual assessments of tree canopy phenology

The leaf phenological stages of each tree within a plot were visually assessed in accordance with established protocols (Liu et al. 2015; Schwartz et al. 2013), with a twice per week observation frequency during the critical phases of bud burst and leaf expansion and a weekly frequency during the other phases (Schwartz et al. 2013; White et al. 2014) (observation dates shown in Figure 1). The visual assessment of leaf phenology started on 3rd March, 2015, in the leaf-off phase of deciduous trees, and extended until after the last sampled tree reached

full leaf/needle expansion (25th June 2015). Dates of three key phenological observation levels (390, 490, 590) were selected to mark the start (SOS), middle (MOS) and end (EOS) of spring season (Schwartz et al. 2013) (remote sensing-derived SOS, MOS and EOS are defined in Figure 2). SOS is associated with >90% of bud open (leaf/candle visible), MOS with >90% of leaf/candle out (not fully unfolded) and EOS with >90% of full leaf unfolded (or needles unfolded from candle). Linear interpolation was used to infer dates of key missing codes. To scale from individual tree phenodates to the plot level (plot level analysis), a plot average was calculated based on a tree observation weighted by its plot percent basal area (White et al. 2014).

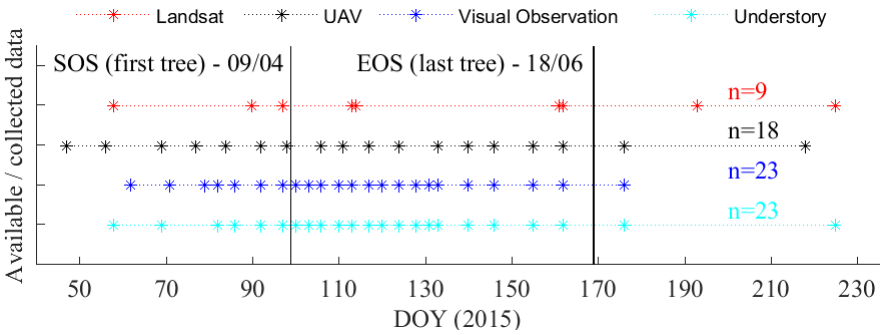

Figure 1. Frequency of data collection/data availability over the study area. The total number of observations are given by 'n'. Vertical lines represent the date when the first tree was observed to reach SOS and the date when the last tree reached EOS. The ground data collection was intensified from ~DOY 90 to ~DOY 130 in order to better monitor SOS.

2.3 Ground photography of understorey

The understorey development within the plots was independently monitored by using a Nikon D300 digital camera (usually on the same dates as the visual assessments, Figure 1). With a Nikon AF NIKKOR 28 mm lens attached to the D300, photography was taken from a fixed position (and viewing angle) slightly outside the plot`s boundary in order to include as

much of the plot`s understorey vegetation as possible within the lens` field of view (example images in Figure S1). Images were then acquired with ISO-400, f/2.8, focus to infinity, auto exposure time and saved in RAW format.

The RAW images were converted into 8 bit TIFF file format and regions of interest (ROI) were defined within each image that maximized the area of understorey for each plot. Average red-green-blue (RGB) digital number (DN) values were extracted from within the images' ROI across all dates and used to calculate time series of Green Chromatic Coordinate (GCC) (Eq. 1) (Sonnentag et al. 2012). Similarly, GCC time series were extracted for a calibration board placed vertically within the camera`s field of view (Figure S1) in order to monitor the sensitivity of GCC values to changes in illumination conditions. The temporal stability of GCC for this board suggests the ability to use it for seasonal understorey dynamics (Figure S2).

$$GCC = \frac{G}{G + R + B} \quad (1)$$

2.4 UAV data collection and processing

The collection, processing, calibration and validation of the series of UAV data used in this paper is described in detail in Berra et al. (2017). Briefly, the study area was flown by one of two fixed-wing UAVs (Quest300 and QPOD - QuestUAV Ltd., Amble, UK) equipped with one unaltered (VIS) and one near infrared (NIR)-modified commercial off-the-shelf (COTS) digital camera. Radiometrically calibrated orthomosaics (5 cm spatial resolution; geolocation accuracy of ±11 cm) were generated for 18 dates (Figure 1), from which DNs were corrected to surface reflectance, based on calibration boards in the imagery (empirical line method (Smith and Milton 1999)), and to Normalized Difference Vegetation Index (NDVI) (Eq. 2).

Orthomosaics were created individually per date and per camera using the software Agisoft PhotoScan v.2 (Agisoft LLC, St. Petersburg, Russia). We found that, at a Landsat 30 m grid scale, consistent NDVI time series, derived from single calibration equations (i.e. derived from a single reference date rather than from every acquisition date) (R^2=0.95 compared to field spectrometer; and R^2=0.88 compared to Landsat 8 data), can be acquired in very variable illumination conditions (Berra et al. 2017), and this dataset is used in this study. Time series of UAV GCC (Eq. 1) were also used, as we showed that this index can appropriately track the phenology of this woodland (Berra et al. 2016). Using NDVI provides a unique opportunity to investigate whether NIR information from COTS sensors can allow a better understanding of vegetation seasonality, as suggested by Brown et al. (2017).

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (2)$$

where ρ_{red} and ρ_{NIR} are reflected light in the red (red channel of UAV VIS camera) and NIR (blue channel of UAV NIR camera) bands, respectively. The UAV NIR camera cannot be used alone to calculate NDVI as the addition of an external long-pass filter excludes wavelengths below 660 nm (the red-edge region); the resulting channels are also highly overlapped (Berra et al. 2017).

To calculate phenology time series from the UAV data, ROIs were defined on the orthomosaics at a tree, plot and Landsat grid scale. While a comprehensive validation of the UAV NDVI time series at a Landsat 30 m scale was performed in our previous study (Berra et al. 2017), the potential of the UAV NDVI time series at smaller spatial scales (tree crown) is investigated in this work. Firstly, in order to assess the potential and accuracy of UAV data for tracking individual-tree level phenology, it is necessary to identify the tree crown boundaries within the validation plots. For this, the location of each sampled tree was overlain on the UAV

255 orthomosaics and each tree crown boundary was manually drawn based on leaf-on
256 orthomosaics.

257 UAV-derived phenometrics (refer to section 2.6 regarding fitting of models) were
258 firstly calculated using time series of vegetation indices (VI) based on the mean DN of all
259 pixels from within a tree crown (as in Klosterman and Richardson (2017)), but this approach
260 proved to be particularly problematic over plots with strong influence of understorey vegetation
261 signal (Figure 2a). The absence of a winter baseline caused the sigmoid model to have a high
262 failure rate in estimating spring phenometrics. For this reason, this study used the mean DN of
263 all pixels with a DN value above the 80th percentile (within each tree crown), based on a
264 sensitivity analysis (Figure S3), to calculate GCC (as in Eq. 1) and NDVI (the mean DNs were
265 corrected to surface reflectance - using the single calibration equations - and NDVI was
266 retrieved as in Eq. 2). The 20% brightest pixels from the VIS green band were firstly detected
267 and their location used as a spatial mask to select DN values from the other VIS and NIR bands.

268 It is likely that higher percentiles of DN values prioritize sunlit pixels, diminishing the
269 influence of understorey/shading and resulting in a stronger seasonal signal (Figure 2b). This
270 is due to the high spatial resolution of the UAV orthomosaics (5 cm), which allows tree
271 elements (e.g. branches), background cover and shadows to be discernible (Figure S4). During
272 leaf-off conditions, we observed (via visual inspection) that the 20% brightest pixels (within a
273 tree crown) were usually associated to tree trunk and branches on the orthomosaics (as
274 exemplified in Figure S4), therefore diminishing the influence of the expected early
275 understorey development in the VI time series (Figure 2).

276 Daily weather variation may influence the brightest pixel selection, especially on sunny
277 days, as only one side of the canopy would receive direct sunlight, likely resulting in brighter
278 pixels in this side (but dependent on the sun-canopy-sensor geometry). Nevertheless, the

dominance of diffuse light in most UAV image acquisitions (16/18) will likely minimise these impacts.

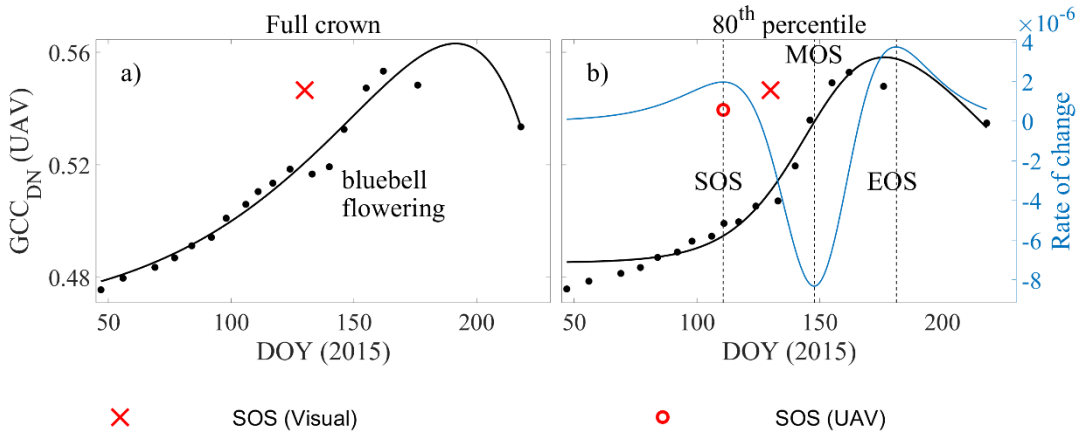

Figure 2. Effect of the 80th percentile method on the smoothed UAV GCC_{DN} time series data (black dots) and consequent phenometric estimations in the Sycamore plot. The data are fitted by the greendown model (solid line), with phenometrics (SOS, MOS and EOS) marked by the vertical dashed lines. The model failed to estimate phenometrics in a). The averaged SOS from visual assessments is also shown for comparison purposes.

Following validation of UAV phenology (see section 2.6), the analysis was expanded over the whole woodland by automating the tree crown detection and delineation (making it easier to implement on an operational basis). Besides providing the opportunity to examine canopy phenology of individual trees over a relatively larger area, this information will also allow quantification of the spatiotemporal variability in leaf phenology within Landsat pixels. The automatic delineation was achieved by using a watershed-based approach, modified from Panagiotidis et al. (2016), who also used very high resolution UAV imagery to delineate tree crowns.

2.4.1 Automatic tree crown delineation

Deriving 3D forest structure from UAV imagery has been possible due to advances in SfM techniques (Snavely et al. 2008), which have allowed Digital Terrain Models (DTMs) and Digital Surface Models (DSMs) to be generated out of a 3D photogrammetric point clouds (St-

Onge et al. 2015). In a forest environment, the underlying ground topography (DTM) can be reconstructed based on points classified as ground (Panagiotidis et al. 2016). A Canopy Height Model (CHM) can therefore be obtained by subtracting these two elevation models and forest attributes (e.g. tree crown area) can be retrieved at an individual tree level (Hernandez et al. 2016). The watershed algorithm is frequently used to perform crown delineation from a CHM (Edson and Wing 2011; Ke and Quackenbush 2011; Mei and Durrieu 2004; Panagiotidis et al. 2016; Zaki et al. 2015).

A flowchart outlines the steps of the automatic delineation method used in this study (Figure S5, with a detailed description in the caption). Briefly, watersheds are extracted from the inverted CHM (via segmentation), so the watershed limits follow gaps between crowns, and, after cleaning (e.g. remove peaks with low height and low brightness) and selection (e.g. select the highest peak within a defined radius), the remaining watersheds represent tree crowns. The quality of watershed-derived tree crowns was evaluated by calculating the producer`s and user`s accuracy (Ke and Quackenbush 2011), using as reference data the manually delineated tree crowns from each plot (as described above). Therefore, individual-tree level phenodates across the entire study area were estimated using the automatically detected tree crowns as ROIs.

2.5 Landsat data

A time series of Landsat 8 Operational Land Imager (L8-OLI) and Landsat 7 Enhanced Thematic Mapper Plus (L7-ETM+) atmospherically corrected surface reflectance images were obtained (processed according to (USGS 2016a, b)) over the study area from January to September 2015 (1 month before and 1 month after the UAV flights). 9 out of 68 images were

selected (Figure 1), following exclusion of cloud/cloud shadow contaminated images (with aid of quality flags) and ETM+ images with significant data gaps (Scan Line Corrector-off) over the study area. Besides NDVI (Eq. 2), GCC was calculated similarly to Eq. 1 but using Landsat spectral reflectance rather than DNs.

The Landsat series, unevenly distributed in time and with a temporal gap of 48 days (DOY 113-161), is acknowledged to be not ideal to consistently estimate the spring phenometric dates of this woodland, but it is representative of the time series achievable with medium resolution satellite sensors in cloudy regions such as the UK (Armitage et al. 2013). Nevertheless, this satellite series can still track the overall dynamic of this woodland and will provide the context and/or data to: i) achieve objective 5 of this research (i.e. assessing spatial variability in phenology at a Landsat sub-pixel level); ii) investigate the effects of two different plant functional types (deciduous and evergreen) in the phenology detection; iii) allow a direct comparison between UAV and Landsat VIs time series (rather than phenodates), therefore showing the quality of time series data from UAV COTS cameras; iv) aggregate the UAV data up to the Landsat scale, allowing for the differences between leaf canopy phenology and LSP to be investigated. However, only Landsat SOS will be compared with UAV phenometrics, as the data gap impedes any meaningful estimate of Landsat MOS and EOS (as seen in Figure 4).

2.6 Analysis methods

A second-order Savitzky-Golay filter (window size = 5) was applied to the original time series data (ground photography, UAV and Landsat) in order to diminish the influence of noise in the series (Lhermitte et al. 2011; Miao et al. 2013; Ryan et al. 2012). Key spring phenological markers were extracted from the smoothed time series of remote sensing products by means of

346 curve fitting. We tested three sigmoid-based models, similarly to Klosterman et al. (2014),
347 namely: the simple (Zhang et al. 2003), greendown (Elmore et al. 2012) and generalized
348 (Klosterman et al. 2014) model. The phenological transition dates were identified by using
349 local extremes in the rate of change in the curvature of the fitted models (Klosterman et al.
350 2014; Zhang et al. 2003). Transitions dates correspond to the times at which the rate of change
351 in curvature exhibits local minima or maxima (as exemplified in Figure 2b). These extremes in
352 the rate of change were associated with start (SOS), middle (MOS) and end (EOS) of spring
353 season, similarly to Klosterman et al. (2014).

354 Residuals from curve fitting (Filippa et al. 2016) were used to calculate measures of the
355 statistical uncertainty in SOS, MOS and EOS dates, for each of the remote sensing products
356 separately, providing the best phenology extraction method for each plot (a detailed description
357 is in the caption of Table S1). However, because the vegetation types outside the plots were
358 unknown, the best general method across all plots was identified in order to allow phenology
359 monitoring at a woodland scale. The uncertainty of the models can be seen in the supplementary
360 material (Table S1 to Table S4), with selected models indicated in the captions of figures and
361 tables throughout this paper.

362 To evaluate the agreement between UAV-derived phenometric dates versus those
363 obtained from each of the independent data sets (visual assessments, ground photography and
364 Landsat), several common measures of statistical agreement were used, including the root-
365 mean-square-error (RMSE), coefficient of determination (R^2) and bias for each case. Because
366 the dependent and independent observations both include significant measurement uncertainty
367 and because the bivariate relationship is symmetric (i.e., the interpretation of the data does not
368 change when the variables assigned to *x*- and *y-axis* are reversed), Reduced Major Axis (RMA)
369 regression (Smith 2009) was used to estimate the slope and intercept (95% confidence interval)

370 on linear regressions between each independent data set and the UAV-derived phenophases
371 dates. Among the three phenometrics, the discussion will focus on SOS, as this transition date
372 is usually of greatest interest in phenology studies.

373 The analysis was firstly carried out at an individual tree level (6 plots, 120 individuals)
374 where UAV derived estimates were compared against visual assessments, providing measures
375 of how well the UAV data can predict leaf phenology event dates. The analysis scaled to a plot
376 level, where ground (visual assessment and understorey development), UAV and Landsat
377 metrics were compared, contributing towards a better understanding of the phenology detected
378 by the UAV sensors. For the UAV data, phenometrics were derived considering the plot
379 boundaries (as seen in Figure 5a) as ROIs (the 80% percentile within each ROI was applied,
380 therefore minimizing understorey effects). These ROIs were thereafter used to weight the
381 Landsat NDVI and GCC values, as the plots were intersected by more than one Landsat pixel,
382 and the weighted time series used to estimate phenometrics.

383 UAV-derived individual-tree level phenodates were estimated across the entire study
384 area using the automatically detected tree crowns (section 2.4) as ROIs. Boxplot statistics of
385 these phenodates were generated by grouping the main tree species into four land covers
386 (broadleaf deciduous, Sitka spruce, Norway spruce and Larch; Figure S6). Besides allowing a
387 detailed phenology map to be produced, UAV estimates were compared against Landsat land
388 surface phenology (LSP) in order to understand leaf phenology variability within Landsat
389 pixels. Two approaches were used to scale from UAV to the Landsat pixel level: 1) a mean
390 UAV phenodate was calculated based on individual-tree phenodates weighted by the percent
391 crown area within each Landsat pixel area, an approach which is similar to the percent basal
392 area (White et al. 2014); and 2) mean UAV DNs were extracted from within each Landsat grid
393 cell, the values of which were converted to VIs in order to estimate UAV phenodates in a

manner more similar to Landsat data, i.e., considering a continuous landscape (30 x 30 m) instead of isolated tree crowns. The latter approach also allowed the dynamics of different UAV- and Landsat-based VIs time series to be compared, rather than phenodates, providing a means to quantify the strength of the relationship between the two datasets. For each available Landsat date (pixel-wise), the closest UAV acquisition date was selected to compose a pair of observations, resulting in between 7 and 9 pairs of observations over the woodland.

3. Results

3.1 Comparison of visual assessments to estimates from UAV time-series data

Substantial phenological variation was detected between individuals of the same species within some plots. Larch presented the smallest variation in the observed key phenodates (7 - 22 days), whilst Sycamore showed the largest variation (29 - 43 days) (Table 1). In general, even when a plot can be considered as having reached a defined phenophase (given by the plot mean), around one quarter of individuals in a plot still lagged considerably. The plot averages show Larch as being the first species to start the growing season on DOY 101 (April 11, 2015) and, 40 days later, Norway spruce as the last one (May 21, 2015) (Table 1). Likewise, Larch was the first species to reach EOS on DOY 118 (April 28, 2015), and, 46 days later Norway spruce reached this same level (June 13, 2015). A comparison between observed phenology dates and DBH and total height did not reveal any consistent relationships ($R^2<0.2$) that were not potentially attributable to interspecific differences (i.e. six plots grouped together, n=120). No significant relationship ($p<0.5$) was observed when the analysis was constrained to within each plot (n=20) ($0 \leq R^2 \leq 0.18$); intraspecific differences may still occur but they are not correlated with DBH or height.

Table 1. *In situ* phenology record analysis: First, average (in bold) and last day of year (DOY) when SOS, MOS and EOS were observed. The 'Variability' shows the difference between the dates that the first and last trees reached each phenophase and the number of trees (in brackets) which reached these phenophases outside the mean ± one standard deviation).

Plot	*N*	Phenophases			Variability		
		SOS	MOS	EOS	SOS	MOS	EOS
Larch	20	099, **101**, 106	103, **107**, 119	106, **118**, 128	07 (*3*)	16 (*3*)	22 (*5*)
Sycamore	20	105, **129**, 143	108, **135**, 151	122, **143**, 151	38 (*6*)	43 (*6*)	29 (*4*)
Oak	20	106, **116**, 124	113, **125**, 137	117, **132**, 140	18 (*4*)	24 (*6*)	23 (*8*)
Sitka s.	20	122, **135**, 146	140, **149**, 159	151, **159**, 169	24 (*5*)	19 (*6*)	18 (*5*)
Norway s.	20	132, **141**, 146	151, **158**, 169	159, **164**, 169	14 (*11*)	18 (*3*)	10 (*7*)
Mix	20	115, **119**, 137	115, **127**, 143	119, **133**, 143	22 (*2*)	28 (*6*)	24 (*4*)
Average	120	113, **124**, 134	122, **134**, 146	129, **142**, 150	21 (*5*)	25 (*5*)	21 (*6*)

We compared these direct observations of canopy phenology with individual tree-level transition dates derived from UAV phenology curves. Between the two UAV-derived indices, GCC-estimated phenodates were consistently most closely associated with the visual assessment of canopy phenology ($0.60 \leq R^2 \leq 0.83$), with the best match occurring with SOS dates (RMSE=8 days) and the weakest relationship with EOS dates (RMSE=13 days) (Figure 3a-c; Table 2). UAV-GCC estimates were biased at 1 (SOS), 5 (MOS) and 10 (EOS) days later with respect to visual assessment, indicating that as spring progressed, the UAV-GCC estimates were increasingly later in comparison to the leaf phenology observations. In terms of SOS, the RMSE value has the same magnitude as the approximate temporal resolution of the UAV data acquisitions (8 days, Figure 3a; Table 2). Within this RMSE is also included the uncertainty of the visual observations, which it was not possible to quantify, but it could be reasonable to assume errors of around 3.5 days (revisit frequency of the visual assessments).

UAV GCC_{DN} can therefore be assumed to estimate SOS dates at tree individual level with an accuracy better than 1 week.

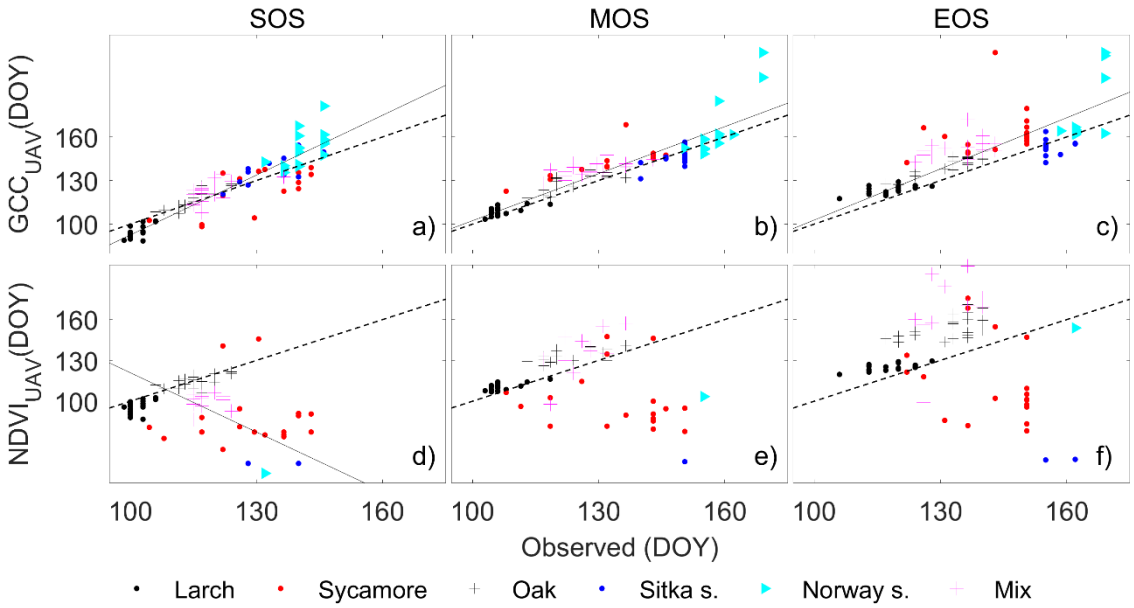

Figure 3. Visually assessed dates compared against dates estimated from the UAV remote sensing derived products, at an individual-tree level. UAV phenometrics are calculated based on the least uncertain sigmoid model per plot (Table S1; Sycamore, Sitka s., Norway s. and Mix = Greendown model; Larch and Oak = Simple model). Dashed lines represent the 1:1 line and solid lines are RMA regression models. Statistics are given in Table 2.

Table 2. Statistics from the comparisons in Figure 3. $^{**}p<0.001$, $^{*}p<0.05$; N is sample size. Bias refers to the average difference between UAV and visual observations.

		RMSE (days)	R^2	Bias (days)	N
Figure 3	a)	8	0.83**	1	106
	b)	9	0.82**	5	106
	c)	13	0.60**	10	106
	d)	22	0.06*	-18	72
	e)	29	0.02	-6	72
	f)	41	0.02	4	72

Phenodates derived from the UAV NDVI presented large mismatches (RMSE≤41 days) and poor correlations (R^2≤0.06) with the visually assessed dates (Figure 3d-f; Table 2). UAV NDVI was also less successful (in terms of function convergence) in extracting phenodates (N=72) from the time series of data than GCC (N=106) (Figure 3). This

result is due to the presence of noise in some of the individual tree time series (UAV data), which contributes towards a weak seasonal signal and poor data quality (at least at an individual tree level), being particularly critical across the evergreen plots (e.g. Figure S7). The negative correlation (R^2=0.06, Figure 3d) is caused by the very early UAV SOS detected for some trees within the Sitka (two trees) and Norway (one tree) spruce plots, reflecting the effect of the poor UAV NDVI data quality (at a tree crown scale) rather than a biological response (Figure S7); when these three points were eliminated (Figure 3d), no significant relationship was observed (R^2=0.01, N=69). For this reason, UAV GCC time series were selected to estimate individual tree-level phenology across the whole woodland (section 3.3).

The comparison between UAV and visual observations across all trees grouped together (Figure 3a-c) suggest that higher uncertainties occur within the evergreen species. The data points related to the Norway spruce plot are the most outlying and the highest failure rate in estimating spring phenometrics occurred within the Sitka spruce plot (*N*=13).

3.2 Plot level comparison of ground photography, UAV and Landsat phenology

In areas of deciduous woodland, where the understorey vegetation greened up earlier than the overstorey (Sycamore, Oak and Mix; Figure 4), Landsat predicted SOS consistently earlier than visual assessments (bias (NDVI) = - 35 days, bias (GCC) = - 24 days, *n* = 3). This suggests that increases in Landsat NDVI and GCC values from the winter baseline are triggered predominantly by the understorey development. On the other hand, the use of the 20% brightest pixels from UAV orthomosaics to generate a GCC time series of data reduced the influence of understorey vegetation (Sycamore, Oak and Mix; Figure 4), resulting in a closer agreement between UAV and visually observed SOSs (bias = -7 days, *n* = 3). Nevertheless, Landsat and

UAV estimates of SOS were more similar (bias (NDVI) = -5 days, bias (GCC) = -14 days, *n* = 1) over the plot in which the understorey greened up after the dominant trees (Larch; Figure 4), i.e., where the understorey did not contribute to the signal resulting from the early stages of bud/leaf development. Later understorey budburst, although not usual, has also been observed in other temperate forest (Richardson and O'Keefe 2009). The later greening-up of the Larch plot understorey could be explained by the high amount of canopy gaps in the Larch overstorey (even after needle expansion, Figure S1), meaning that high-light conditions were available for a longer period of time on the ground; therefore, not justifying the adoption of a strategy of phenological escape (Richardson and O'Keefe 2009).

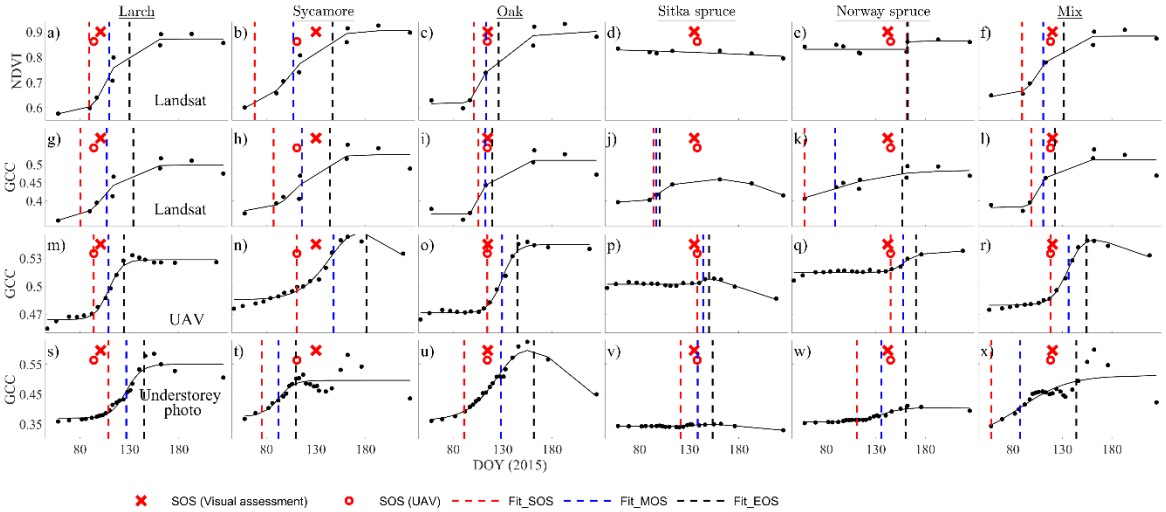

Figure 4. Plot level comparison between Landsat (a-l), UAV (m-r) and understorey ground photography (s-x). The time series of data (black dots) are fitted by the "best model" per plot (Table S1 to Table S4), with phenometrics marked by the vertical dashed lines (preceded by 'Fit_' in the legend). The averaged SOS from visual assessments and the UAV SOS are also shown for comparison purposes. Landsat MOS and EOS metrics are shown in this figure, but are not considered in the comparisons with the UAV phenology (these metrics are likely not meaningful due to the large Landsat data gap from DOY 113-161).

As it would be expected, a weak seasonality was detected by the Landsat NDVI and GCC series over the evergreen covers. Comparatively, a stronger response was detected by the

UAV GCC, mainly over the healthy Norway spruce plot (Figure 4q). A severe defoliation was observed in the Sitka spruce plot, likely caused by an outbreak of *Elatobium abietinum* (green spruce aphid). The effect of such defoliation was detected by the UAV (more evidently) and Landsat remote sensing products as the VIs experienced lower values towards summer.

A deviation from the expected trend occurred in the Sycamore and Mix plot due to bluebell (*Hyacinthoides non-scripta*) flowering, as detected by the understorey photography (Figure 4t,x). The abundant presence of blue flowers in these plots increases the blue channel's DN values on the ground photos, consequently decreasing GCC values temporarily. Such an effect was also detected on the UAV GCC time series when the average of the full crown was tested (Figure 2a), a trend which largely disappeared when the average of the brightest pixels was used (Figure 2b). Therefore, these results suggest that different understorey species can have significant influence on the integrated ecosystem signal, especially with very high spatial resolution optical sensors.

3.3 Individual tree-level phenology across a small woodland

4354 tree crowns were automatically delineated from the UAV orthomosaics with an overall accuracy of 63%, as given by the producer`s (PA) and user`s (UA) accuracy. Conifer species were more accurately delineated with PA ranging from 73% to 80% (UA from 78% to 82%), whilst broadleaf species achieved lower PAs (43% to 57%, with UA from 43% to 63%). These accuracies are within the range reported in the literature (Li et al. 2016; Lim et al. 2015; Panagiotidis et al. 2016; Thiel and Schmullius 2016) and once more indicate that complex forest structures (as in the Oak and Mix plots in this study) are challenging environments for automatic extraction of tree attributes (Duncanson et al. 2014). Despite these uncertainties, the

517 average SOS date estimated from the automatically delineated crowns (126±10 days) was
518 remarkably similar to the reference date from manually delineated crowns (127±10 days). A
519 similar tendency was found for MOS (142±9 days for reference and 141±9 days for automatic)
520 and EOS (155 ±13 days for reference and 155 ±13 days for automatic). This suggests that the
521 automatic delineation errors should have limited impacts on the investigation of the individual
522 tree-level phenology across this woodland (community phenology).

523 Analysis of boxplots (Figure 6) and visual inspection of the date of onset from the
524 UAV-derived map (Figure 5a) show substantial intra- and inter-specific variability in
525 individual-tree level leaf phenology across a 15 ha woodland. The uncertainty of the UAV SOS
526 transitions (<5 days for 86% of the mapped trees, Figure S8) is smaller than the range of dates
527 in Figure 5a, indicating the effectiveness of UAV data for investigating phenological
528 differences among tree species populations and communities. Larch trees leafed out
529 consistently earlier than the others species (Figure 6), as also shown by the blue tones in the
530 eastern part of the woodland (Figure 5a), and present the least intra-specific variation in SOS
531 dates. The Larch cover has a significant number of outliers (Figure 6) which is likely due to
532 the presence of Sitka spruce trees within the defined limit of this land cover, as Sitka spruce
533 leafed out consistently later than Larch (Table 1); this will reflect the contrasting blue and red
534 tones of some adjacent trees in the southern area near the Sycamore plot (Figure 5a).

535 At the other extreme, later onset dates consistently occurred in the Norway spruce
536 compartment, as quantified in Figure 6 and showed by the yellow-red tones in Figure 5a.
537 Intermediate dates of SOS occurred for the deciduous broadleaf species (as sampled in the
538 Sycamore, Oak and Mix plots). Sitka spruce started the growing season, in general, after the
539 broadleaf deciduous trees but before Norway spruce. These spatio-temporal patterns of onset

dates show a correlation with tree species communities, matching the chronological order of visually assessed leaf phenology events within the six plots (Table 1).

Across the Sitka spruce area, a considerable number of trees have either no (Figure 5a) or highly uncertain (Figure S8a) SOS estimates. This could be due to a weak seasonal signal, reflecting the defoliation caused by the aphid outbreak (as observed in the plot level analysis, Figure 4p), which may have attacked some trees more severely, impeding needle unfolding and consequently limiting greening up and significant increases of UAV GCC_{DN} values. Overall, missing phenometric retrievals can be explained by the failure of the fitting function to converge because of poor data quality (e.g. noise and data gaps) or weak seasonal signal (D'Odorico et al. 2015).

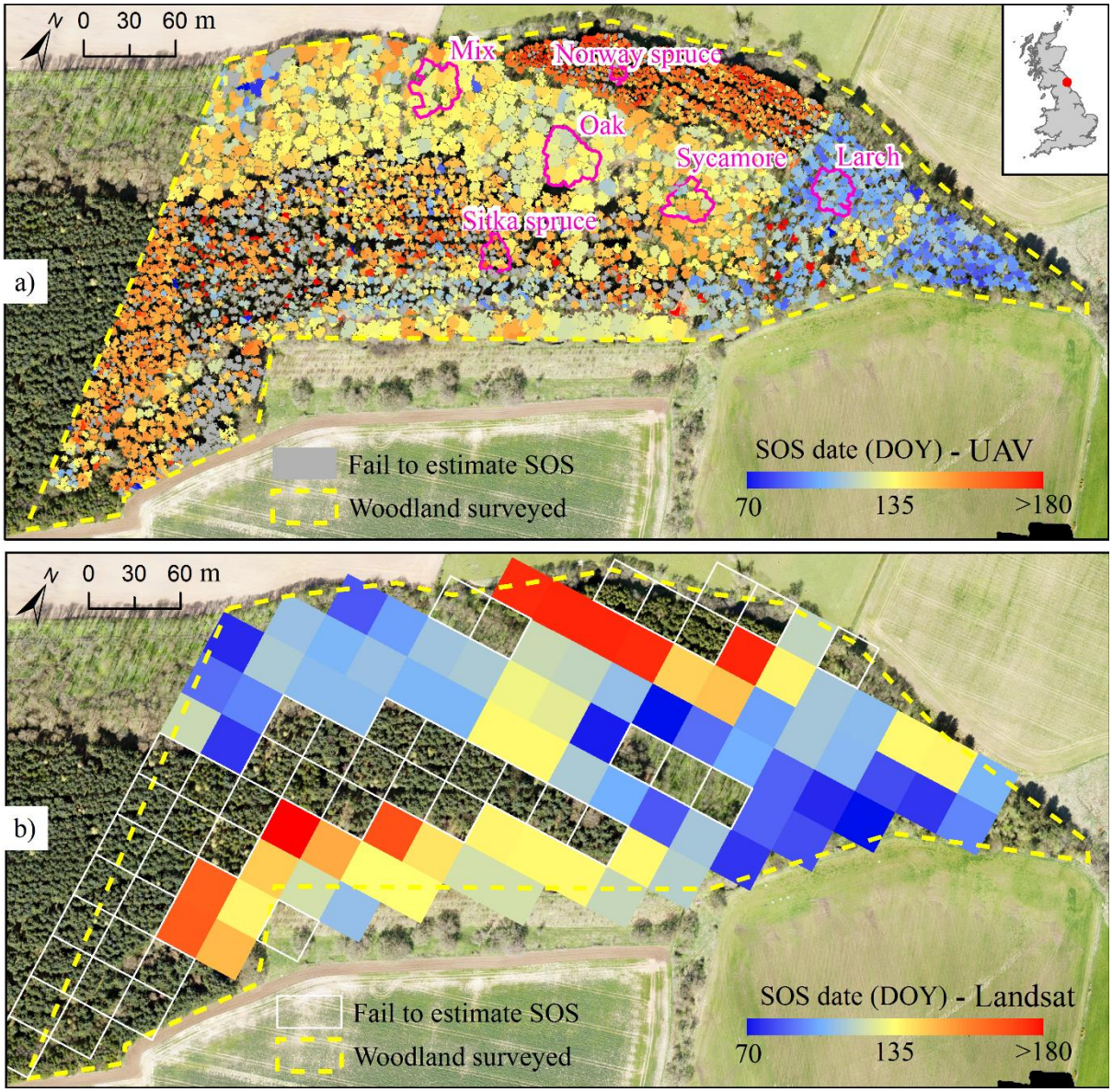

Figure 5. a) Location of the study area and individual tree-level predictions of SOS using UAV GCC_{DN} (80^{th} percentile; fitted by the greendown model), and b) pixel-level predictions of SOS using Landsat NDVI (fitted by the simple model). Background orthomosaic derived from UAV images (visible camera) acquired on 21/04/2015 (DOY 111).

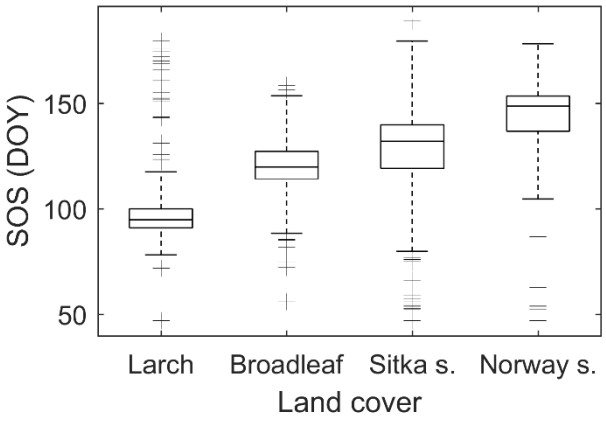

Figure 6. Boxplots of SOS dates of individual trees (n = 4354), as mapped in Figure 5a. For each boxplot, the central mark represents the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers (<2 times the standard deviation). Broadleaf encompasses the Sycamore, Oak and Mix plots.

3.4 Comparing UAV and Landsat phenology at different scales

A SOS map was also generated based on Landsat NDVI time series (Figure 5b) (an uncertainty map can be seen in Figure S8b). Despite species-related patterns being less clearly depicted (in comparison to the UAV individual tree-level map), some broader inferences can be drawn. Visual assessments (Table 1) and the UAV SOS map (Figure 5a) showed Larch as the first tree species to start the spring season, a dynamic which was not observed on the Landsat SOS map (Figure 5b). Instead, areas in and around the Sycamore and Mix plots had the earliest SOS dates, which could be due to the early understorey development, as noted in the plot-level analysis (Figure 4). Either no (Figure 5b) or highly uncertain (Figure S8b) SOS estimates were, generally, observed over the Sitka spruce area, which is due to the combined effects of weak seasonal signal and the aphid outbreak. Landsat pixels intersecting Oak and Larch plots seem to reproduce more closely the phenology patterns observed at the UAV canopy level.

Linking such integrated landscape processes (Figure 5b) with fine scale information (Figure 5a) can therefore be challenging. Large temporal variations in averaged individual tree-

level canopy phenology occurred within Landsat pixel areas across the woodland (RMSE=32 days, R^2=0.11; Figure 7b), even after mixed and evergreen land covers were masked out (RMSE=19 days, R^2=0.01; Figure 7a). Comparisons with Landsat GCC produced similar results, which can be seen in Figure S9. This suggests that the average phenology of the dominant canopies (as mapped by UAV data) may lag considerably from the Landsat LSP, as also detected in the plot-level analysis (Figure 4); or, that such differences could be due to the low quality Landsat time series.

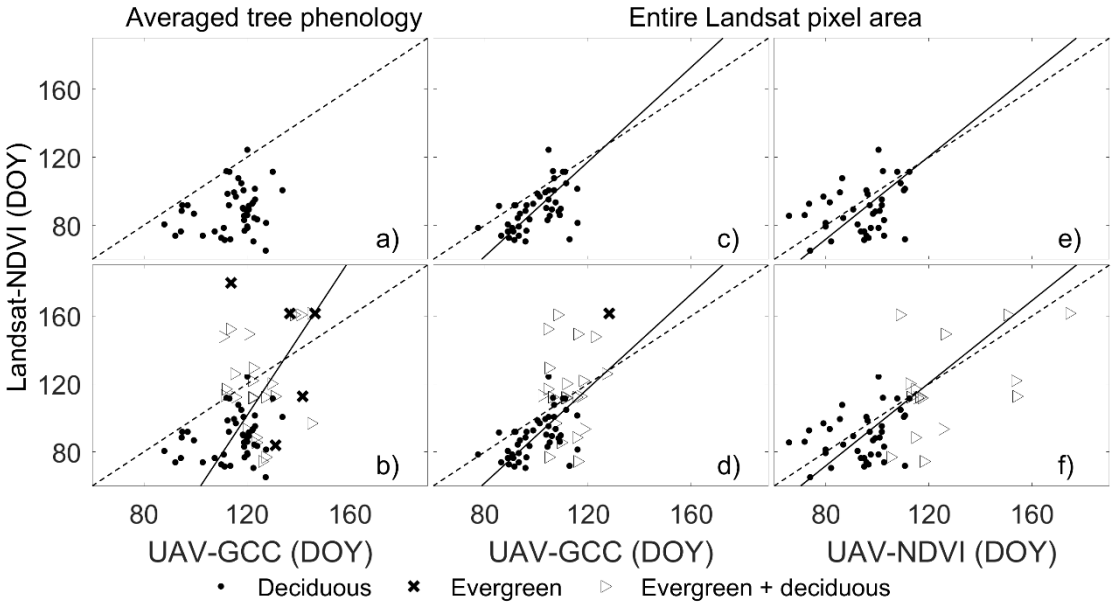

Figure 7. Comparison between UAV SOS and Landsat SOS. Landsat LSP (simple model) compared with: a-b) UAV-derived (greendown model) averaged tree-level phenology (within each Landsat pixel area); and c-f) with UAV LSP (greendown model) considering the entire Landsat pixel area. Pure deciduous covers are shown separately (top row). Dashed lines represent the 1:1 line and solid lines are RMA regression models (**p<0.001; *p<0.05). Statistics are given in Table 3.

Table 3. Statistics from the comparisons in Figure 7. **p<0.001, *p<0.05; N is sample size Bias is calculated relative to UAV, so a negative bias indicates that the corresponding Landsat estimate is earlier.

		RMSE (days)	R^2	Bias (days)	N
Figure 7	a)	19	0.01	-28	47
	b)	32	0.11*	-19	76
	c)	12	0.28**	-11	44
	d)	21	0.32**	-6	67
	e)	17	0.12*	-5	41
	f)	25	0.22**	-5	59

UAV orthomosaic DNs were aggregated up to the Landsat scale, i.e., the ROI was the entire Landsat pixel area rather than just tree crowns. Firstly, a direct comparison between UAV and Landsat VIs time series (rather than phenodates) was undertaken. This comparison showed varying degrees of correspondence depending on the pair of VIs considered ($0.45 \leq R^2 \leq 0.91$, Figure 8), where UAV NDVI had the best overall agreement with Landsat NDVI, especially over deciduous-only areas ($R^2=0.91$; Figure 8h). Despite these moderate/strong relationships in spectral information (Figure 8), distinct differences in the dates of phenological events predicted by the different UAV and Landsat VIs are clearly evident (Figure 7c-f), with UAV data explaining less than 50% of the variability in Landsat-based estimates of SOS dates in any of the comparison cases (comparisons with Landsat GCC produced similar results, Figure S9). Following exclusion of evergreen covers, the average RMSE improved from 23 (Figure 7d,f) to 14 (Figure 7c,e) days; Landsat GCC agrees best with UAV GCC (RMSE=10 days, $R^2=0.5$; Figure S9c). Therefore, spatial averaging of UAV data and exclusion of evergreen areas improves the agreement with Landsat phenology, but significant discrepancies remain.

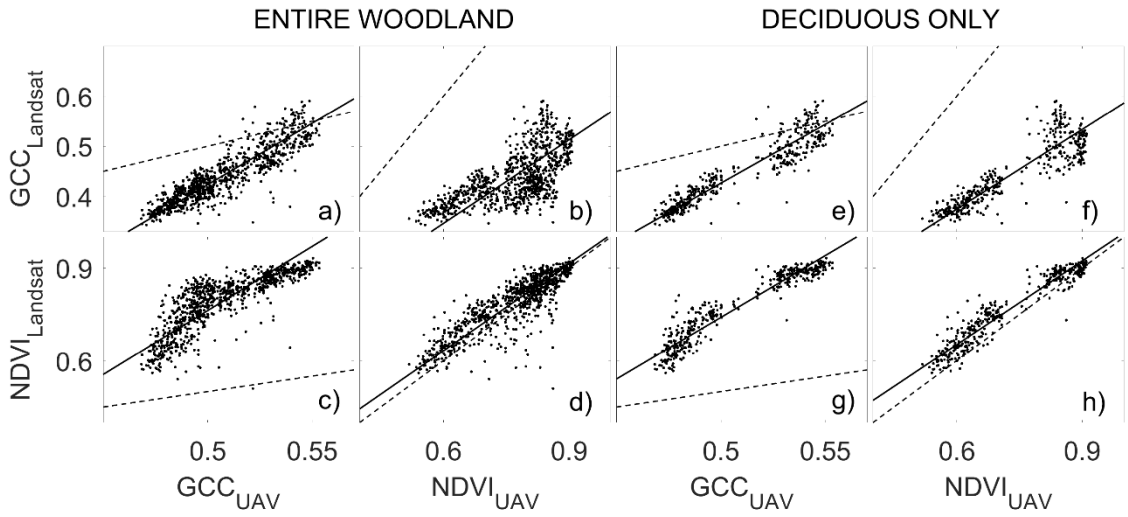

Figure 8. Comparison between UAV- and Landsat-based VI time series, at a Landsat pixel scale, considering the entire woodland (a-d) and only the deciduous cover (e-h). UAV VIs are based on the mean of the whole polygon (Landsat grid). Dashed lines represent the 1:1 line and solid lines are RMA regression models. Statistics are given in Table 4.

Table 4. Statistics from the comparisons in Figure 8. **p<0.001. N is sample size.

		R^2	Bias	N
Figure 8	a)	0.80**	-0.07	878
	b)	0.45**	-0.33	878
	c)	0.66**	0.28	878
	d)	0.74**	0.02	878
	e)	0.87**	-0.06	401
	f)	0.79**	-0.29	401
	g)	0.90**	0.27	401
	h)	0.91**	0.04	401

4. Discussion

4.1 Individual tree level phenology across a woodland

Studies using remote sensing to monitor forest phenology have used, until now, medium to coarse resolution imagery from sensors such as Landsat, MODIS, AVHRR and SPOT-Vegetation, which allow regional to global patterns to be observed but cannot resolve

636 species-scale seasonal dynamics (Fisher and Mustard 2007; Melaas et al. 2016; Miao et al.
637 2013; White et al. 2009). In this study, an effective approach for mapping phenology of
638 overstorey vegetation at a detailed biological scale and across local spatial extents was
639 proposed by using time series of UAV remotely sensed data, offering an intermediate level of
640 observation between ground/near-surface and orbital scales.

641 Visual assessments of species-specific individual tree phenology revealed a spatially
642 heterogeneous ecosystem, where intra- and interspecific differences were up to ~40 days. This
643 unequal budburst and leaf/needle development agrees with observations across other temperate
644 forests (Schwartz et al. 2013; White et al. 2014) and can be due to several factors acting
645 together, such as hereditary influences (Morin et al. 2010) and site-specific factors (Fisher et
646 al. 2006; Ibáñez et al. 2010). Such large variations in timing of leafing-out, despite the trees
647 growing in close proximity (10-30 m), may have implications for studies using ground and
648 near-surface sensors, viewing a small number of trees, to validate satellite-based phenometrics
649 or to characterize the phenology of a population. There are also implications for coarse spatial
650 resolution satellite data, as there is an intrinsic averaging and missing of potential to distinguish
651 these differences in phenology within the pixel area.

652 The within-plot variability in canopy phenology, as assessed by ground observations,
653 was consistently detected by the UAV dataset, providing a new methodological approach to
654 track phenological dynamics within plant communities. An individual tree level detection of
655 spring phenological transition dates was possible due to the user-defined temporal (~7 days)
656 and spatial (5 cm) resolution with which the UAV data was acquired in this research. Overall,
657 UAV-derived SOS could be predicted more accurately than MOS and EOS, with an accuracy
658 of 8 days across the six plots (six tree species). Because the validation was based on visual
659 assessments, which itself is not free of uncertainties (Klosterman et al. 2014; Schaber 2002), it

660 would be reasonable to expect higher accuracies from these UAV estimates. The work of
661 Klosterman and Richardson (2017) suggest that higher accuracies can be achieved in the
662 individual tree-level monitoring of UAV SOS (RMSD = 4.7 days, *n*= 30) and MOS (RMSD =
663 3.6 days, *n*=30), which could be due to the higher frequency of UAV data acquisitions in their
664 study (~5 days).

665 Ideally, comparing ground phenology observations of trees to LSP-derived
666 phenometrics entails the observed trees being representative of the tree species within the
667 pixel(s) area (Delbart et al. 2015; Liang et al. 2011), a requirement seldom met due to
668 difficulties in sampling large areas or the majority of the species, leading to imprecise
669 characterization of species-level phenology (Klosterman et al. 2018). Phenological maps of
670 individual trees, such as the ones generated across the 15 ha woodland in this study (Figure 5),
671 meet this need and represent a powerful tool to visually represent spatio-temporal patterns in
672 canopy phenology, in complement to quantitative analysis. Such georeferenced information
673 potentially brings the opportunity to track the phenological behaviour of all individual trees in
674 the upper canopy (dependent upon a robust automatic tree crown delineation approach),
675 therefore it is possible to census the plant phenological behaviour within a local area or satellite
676 footprint. This, in turn, can be particularly useful to adequately represent variability of highly
677 heterogeneous ecosystems (Fisher and Mustard 2007).

678 An individual-tree SOS map clearly showed the spatial variance in leaf phenology of
679 this woodland, demonstrating how contrasting canopy phenological events can be within
680 Landsat pixels. Therefore, heterogeneity in canopy phenology can still be an issue even at the
681 spatial resolution of Landsat, agreeing with other studies (Fisher et al. 2006; Klosterman et al.
682 2018; Liu et al. 2017). This demonstrates that caution should be taken when using Landsat to
683 understand coarser spatial resolution LSP, based on the assumption that the vegetation

684 phenology within the 30 m scale would be relatively homogeneous (Fisher and Mustard 2007;
685 Liu et al. 2017). It also confirms suggestions that high plant species diversity and pronounced
686 heterogeneity in timing of phenological events have a large influence on the accuracy of LSP-
687 derived SOS (Delbart et al. 2015; Fisher et al. 2006; Hufkens et al. 2012; Melaas et al. 2016;
688 Polgar and Primack 2011).

689 UAV data, as presented in this study, could potentially contribute to phenology studies
690 by providing local-scale measurements meeting two research needs: 1) at the plant- and plot-
691 based scale (Hunter and Lechowicz 1992; Lechowicz 1984; Schaber and Badeck 2003) and 2)
692 continuously in space, similar to remote sensing satellites (Fisher et al. 2006; White et al. 2014).
693 This multi-scale information can provide a comprehensive insight into how the ecosystem is
694 functioning, being useful for many ecological/phenological applications, such as investigation
695 of spatial scaling effects on LSP (Klosterman et al. 2018; Peng et al. 2017), tree species or plant
696 communities mapping (Lisein et al. 2015; Michez et al. 2016), validation and refinement of
697 plant phenological models (Chuine et al. 2013), modelling seasonal carbon sequestration
698 (Wilkinson et al. 2012), understanding the chronological order of phenological events among
699 individual trees (Delpierre et al. 2017) and monitoring the length of the growing season (Norby
700 et al. 2003). Within a climate change context, monitoring techniques which are able to capture
701 the individual phenology of the plants are needed, as not all species are responding similarly
702 (Ibáñez et al. 2010; Thompson and Clark 2008; Vitasse et al. 2009). Furthermore, it is vital to
703 observe a large number of individuals within a forest community in order to adequately capture
704 the expected large variation in timing of phenological events (Donnelly et al. 2017).

705 Despite the broad and ever-increasing use of phenocams (Brown et al. 2016), there are
706 still uncertainties related to the effect of their oblique viewing angles on the temporal trajectory
707 of canopy greenness (Keenan et al. 2014). Furthermore, inclined phenocams can only monitor

708 one side of a tree crown, each of which has a different sensor-target distance; this can impact
709 on data quality as further targets will have weaker signals and stronger atmospheric influence
710 (Richardson et al. 2009). A UAV data set, preferably acquired using the same camera as the
711 phenocam, could provide an insight into whether these issues have significant impacts on the
712 phenocam-derived phenological signal. One of the ultimate applications of UAV-derived
713 phenology would be the validation of satellite LSP products at an appropriate spatiotemporal
714 scale. Because the UAV-derived information is synoptic and georeferenced, comparisons with
715 satellite can be done very precisely, across areas (local scale) comparable with medium and
716 moderate spatial resolution satellite pixels (e.g. Landsat, MODIS). Furthermore, the ability to
717 monitor species-specific phenology with UAVs (although an accurate species identification
718 still requires field observation) could help in understanding the effects of the underlying spatial
719 complexity and heterogeneity of the objects which might be present within the resolved satellite
720 scene (White et al. 2009). For example, the effects of the background cover, which can bias
721 and even hamper LSP transition date estimates (Delbart et al. 2005; White et al. 2009), could
722 be better understood with UAV data, as this can allow tracking of canopy phenology, as
723 opposed to ecosystem dynamics detected by satellites. This would provide new insights into
724 the ecological and biophysical meaning of the spectral information within the satellite pixels.

725 The findings of this research support the use of UAV data for this application, as canopy
726 spring phenology of individual trees (based on the brightest pixels within a tree crown to
727 diminish uncertainties caused by understorey contribution) were consistently detected across a
728 15 ha woodland. Additionally, very good relationships between UAV NDVI and Landsat
729 NDVI were observed (R^2>0.7), especially over deciduous areas (R^2>0.9) (Figure 8), mirroring
730 *in situ* (radiometer) and MODIS NDVI comparisons (R^2=0.92) carried out by Hmimina et al.
731 (2013). Nevertheless, poor relationships were found between UAV- and Landsat-derived SOS
732 in this study (R^2<0.5), indicating that further work is needed to fully assess the potential of

733 UAV data sets for satellite phenology validation purposes due to, mainly, the temporal gaps
734 present in the Landsat data for this site.

735 **4.2 Considerations for use of UAV measurements**

736 Previous studies using near-surface GCC from COTS cameras noted that the
737 relationship between this index and commonly used satellite VIs (e.g. NDVI) tend to saturate
738 at medium/high values of satellite VI, as GCC is less sensitive to increases in LAI due to the
739 lack of a NIR band (Brown et al. 2017; Keenan et al. 2014). It is suggested therefore that the
740 use of NIR band in the near-surface VIs should partly overcome this challenge over areas with
741 medium to high biomass (Brown et al. 2017). Although times series of UAV NDVI could be
742 retrieved with confidence at a 30 m Landsat scale (Figure 8d,h and Berra et al. (2017)), GCC
743 proved to be advantageous at a tree crown scale, as higher quality time series were retrieved
744 and GCC-estimated phenodates were consistently more closely associated with the visual
745 assessments of spring canopy phenology (Figure 3). It is inferred that, at small spatial scales,
746 GCC is better able to take into account the different illumination conditions experienced on
747 some acquisition dates than NDVI.

748 Because GCC uses only data from a single camera (VIS), the three RGB channels
749 should be affected similarly by the varying illumination conditions (as exemplified in Figure
750 S10), independent of the spatial scale in which the data are aggregated. On the other hand, the
751 VIS (unmodified camera) and NIR (modified camera) orthomosaics` DNs might be affected
752 differently by varying illumination conditions over small areas (but not significantly at a 30 m
753 Landsat scale (Berra et al. 2017)). This is due to short time lags between the original VIS and
754 NIR single images automatically chosen to compose the orthomosaics in such areas, hampering
755 the ability of NDVI to normalize for this effect. This time lag could be due to one of the cameras

failing to record an image at a programmed time, while the other does, resulting in pixel values in the final orthomosaic arising from different times within the flight. Viewing angles and location and amount of artifacts (Samiappan et al. 2017) might also be different for the NIR and VIS orthomosaics (as exemplified in Figure S11). Additionally, even though the COTS cameras were flown concurrently, they were located at ~10 cm apart in the UAV frame, resulting therefore in slightly different view angles. Finally, VIS and NIR orthomosaics can be co-registered with an accuracy of ±11 cm (~2.1 pixels, Berra et al. (2017)), which may influence analysis over smaller crowns. Therefore, the combination of all these factors resulted in a noisier UAV NDVI time series, contributing towards a weak seasonal signal and less accurate phenological date estimates at an individual tree-level scale.

Nevertheless, a better quality time series of UAV spectral VIs could potentially be achieved at small spatial scales. It is possible to use a single modified camera recording visible and NIR wavelengths (Berra et al. 2015; Hunt et al. 2010; Verhoeven 2012). Because the three bands are already registered, NDVI could potentially be calculated more consistently at small spatial scales, similarly to GCC from an unmodified camera. It would be valuable, in a future study, to test a dual-camera system composed of an unmodified COTS camera (for GCC) and a modified camera, as proposed above, specifically for NDVI. Furthermore, non-COTS multispectral sensors, such as MicaSense® (Samiappan et al. 2017), are an option and could be advantageous.

More accurate and precise spring phenological transition dates were detected from UAV time series having well-characterized winter base lines and summer plateaus, as observed in the Larch and Oak plots (Figure 4). Plots with only a few data points to characterize either the winter baseline (e.g. Sycamore) or the summer plateau (e.g. Norway spruce) returned, generally, less certain estimates (Table S1). Particularly, the UAV data gap between DOY 176-

218 (42 days) might have increased uncertainties in the EOS estimates across the plots with later greening up (Sycamore, Sitka spruce and Norway spruce; Table 1) as there were fewer data points to characterize the spring to summer transition (Figure 4n,p,q), i.e., the second maxima in the curvature change rate (Figure 2). This suggests that in order to characterize spring events of different species, it would be beneficial to start the data collection from the start of winter up to the end of the summer (without large temporal gaps), a time span which could allow detection of the winter baseline and the summer plateau of a wide range of species (at least in this ecosystem).

Although UAVs can allow a detailed analysis of spatial and temporal patterns across the landscape, this is only feasible at a local scale. A fixed-wing UAV was used to acquire time series of imagery to monitor phenology of a 15 ha woodland in this study. Using a fixed-wing UAV was only possible as the woodland was surrounded by relatively flat crop fields, providing very good ground conditions for taking-off and landing. Other study areas where open spaces are not available close by, rotary-wing UAVs might be a better (if not the only) choice, but smaller areas are then likely to be surveyed. A challenge therefore remains in measuring larger areas, which could be useful, for example, to investigate the seasonal signal of low spatial resolution imagery (>250 m) or to have a representative sample of a large forest ecosystem. Nevertheless, with the continuous technological advances, civil UAVs are expected to provide longer flying times, which could allow larger areas to be surveyed. However, this benefit can be limited by aviation regulations within each country, such as requirement to retain line-of-sight during operations (Torresan et al. 2017). Finally, whilst multiyear UAV observations are still logistically challenging (mainly due to human resources needed), UAV deployment flexibility can increase the number of study sites that can be observed, providing detailed understanding of phenology in understudied biomes or regions. Independent of the camera and UAV model, a key aspect is acquiring a UAV data set which can allow high quality

time series of orthomosaics to be generated. This can be achieved by acquiring high quality UAV images, with a high overlap (>9 images per point was achieved in all dates in this study) and at varying viewing angles, as these characteristics are beneficial if processing the block of UAV images with a SfM-based software (e.g. PhotoScan and Pix4D). Additionally, because time series data at an individual tree level is needed, an accurate georeferencing approach should be utilised by using either Ground Control Points (GCPs) surveyed with Differential Global Positioning Systems (DGPS) (as in this study) or geotags from the original images (direct georeferencing) if a high accuracy GPS is available on-board the UAV. In fact, direct georereferencing seems to be a critical point towards a more automated data collection and processing, as such technology can eliminate the need for: i) GCPs on every acquisition date, and ii) identification and manual allocation of GCPs on the several images, per flight, per camera and per acquisition date.

4.3 Uncertainties in the UAV vs Landsat comparisons

It is somewhat surprising that the UAV data was able to only explain <50% of the variation in phenodates estimated by the Landsat sensors, as the phenology of (theoretically) every tree within a Landsat pixel was detected by the UAV data. Previous studies have indicated that if the validation dataset is able to account for the spatial heterogeneity in timing of phenological events and heterogeneity in species composition within a pixel area, then low uncertainties in the estimated phenodates could be expected (Liang et al. 2011; Liu et al. 2015). However, a close analysis of the two datasets indicates that several causes may have contributed to the pronounced differences between UAV and Landsat phenology.

Evergreen areas, or areas with a low fraction of deciduous coverage, can add significant uncertainty to phenological predictions from orbital sensors (Hmimina et al. 2013; Klosterman

et al. 2014). In this study area, the inclusion of evergreen areas in the UAV vs Landsat comparisons of SOS dates increased RMSE values to up to two weeks (Figure 7, Figure S9). This could be explained by the subtle seasonal signal produced by evergreen covers, which may not be detected by the Landsat sensors, therefore hampering time series fitting and predictions. In addition, a severe pest attack caused defoliation over the largest evergreen stand (Sitka spruce), further decreasing the signal amplitude of this plant functional type and making it challenging to detect even with UAV sensors.

Spatial misalignment of the different remote sensing data sets and the measurements derived from them may also be a contributing factor (Xin et al. 2013). While UAV orthomosaics are georeferenced with a decimetre accuracy (Berra et al. 2017), Landsat products (Level 1T) are expected to have a planimetric accuracy of ~12 m (Storey et al. 2014). Therefore, comparing a Landsat pixel with the same area of an UAV orthomosaic, one can expect misregistration inconsistencies. Additionally, different sensors can have different data quality (Zhang et al. 2017) and differing spectral bands, which provide sensitivity to different vegetation dynamics (Brown et al. 2017).

Different temporal resolutions between UAV and Landsat datasets are likely to be a major source of uncertainty in our comparisons. The frequency of high-quality satellite observations can have substantial impact on phenological detections during periods of phenological changes (Baumann et al. 2017; Fisher et al. 2006; Zhang 2015; Zhang et al. 2009; Zhang et al. 2017), which could bias estimated spring onset by over one week dependent upon configurations of image availability (Melaas et al. 2013; White et al. 2014). Although representative of typical Landsat cloud-free scene availability in the UK (Armitage et al. 2013), in this research, only 7-9 good quality Landsat images remained after a quality check, unevenly distributed in time and with a temporal gap of 47 days from 24th April to 10th June, 2015, a key

period of vegetation green-up. On the other hand, a consistent temporal resolution of around one observation per week was achieved with the UAV. Further studies are necessary to investigate whether the relationship between UAV- and Landsat-derived phenometrics would actually improve if the two data sets had the same (or very similar) temporal resolution (which potentially could now be achieved if combining Landsat 8 and Sentinel-2 imagery (Jönsson et al. 2018)).

5. Conclusions

An effective approach for mapping phenology of overstorey vegetation at a detailed biological scale and across local spatial extents was proposed in this study by using time series of UAV remotely sensed data. This information could aid in further understanding of phenological triggers and biophysical processes of different plant functional types during the critical time of growing season onset.

We presented here the first phenological map of individual trees of an entire woodland, showing that UAV data has the potential to capture species-specific phenology, as opposed to the overall phenological dynamics recorded by orbital sensors. Nevertheless, satellites remain the only feasible tool for large and global scale monitoring of Earth dynamics (Liu et al. 2017), so it is important to continuously improve and refine our understating of LSP with aid of different multi-scale ground truth data.

Although a dual COTS camera system could produce consistent time series of UAV NDVI at a Landsat 30 m scale, GCC based on uncalibrated DNs proved to be more appropriate to track the canopy phenology of individual trees in this study. This does not necessarily mean

that UAV GCC should be preferred over UAV NDVI, but, rather, it reflects the challenge of using two separate COTS cameras to detect the red and NIR bands used to calculate NDVI of individual trees.

Calculation of GCC simplifies data acquisition and processing and is a measure commonly available from phenocams, but may be less suitable for linking directly with reflectance-based satellite data. Therefore, there is future opportunity to investigate whether NDVI derived from a single modified COTS camera or multispectral sensor would produce robust time series measures at small spatial scales. Continued research with similar techniques would further advance the synergism of multi-scale remote sensing observations of vegetation phenology.

6. Acknowledgement

The work of E. F. Berra was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Brazil), grant 1121/13-8. The authors would like to thank Dr. G. Llewellyn from the Natural Environment Research Council (NERC) Airborne Research and Survey Facility for his assistance with the vignetting correction, Dr. C. MacLellan, Dr. A. MacArthur (from the NERC Field Spectroscopy Facility), and Mr. S. Gibson-Poole for their assistance with the camera spectral sensitivity determination, and Dr. M. Peppa, Dr. M. Smigaj, and Mr. M. Robertson for their substantial help with the UAV data collection. Finally, the authors would like to thank Dr. Angela Erb and Prof. Crystal Schaaf (University of Massachusetts, Boston) for pre-processing and making available the Landsat data used in this study.

7. References

Anderson, K., & Gaston, K.J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment, 11*, 138-146

Armitage, R.P., Alberto Ramirez, F., Mark Danson, F., & Ogunbadewa, E.Y. (2013). Probability of cloud-free observation conditions across Great Britain estimated using MODIS cloud mask. *Remote Sensing Letters, 4*, 427-435

Bartholomé, E., & Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing, 26*, 1959-1977

Baumann, M., Ozdogan, M., Richardson, A.D., & Radeloff, V.C. (2017). Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. *International Journal of Applied Earth Observation and Geoinformation, 54*, 72-83

Berra, E., Gibson-Poole, S., MacArthur, A., Gaulton, R., & Hamilton, A. (2015). Estimation of the spectral sensitivity functions of un-modified and modified commercial off-the-shelf digital cameras to enable their use as a multispectral imaging system for UAVs. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-1/W4*, 207-214

Berra, E.F., Gaulton, R., & Barr, S. (2016). Use of a digital camera onboard a UAV to monitor spring phenology at individual tree level. In, *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3496-3499)

Berra, E.F., Gaulton, R., & Barr, S. (2017). Commercial Off-the-Shelf Digital Cameras on Unmanned Aerial Vehicles for Multitemporal Monitoring of Vegetation Reflectance and NDVI. *IEEE Transactions on Geoscience and Remote Sensing, 55*, 4878-4886

Brown, L.A., Dash, J., Ogutu, B.O., & Richardson, A.D. (2017). On the relationship between continuous measures of canopy greenness derived using near-surface remote sensing and satellite-derived vegetation products. *Agricultural and Forest Meteorology, 247*, 280-292

Brown, T.B., Hultine, K.R., Steltzer, H., Denny, E.G., Denslow, M.W., Granados, J., Henderson, S., Moore, D., Nagai, S., SanClements, M., Sanchez-Azofeifa, A., Sonnentag, O., Tazik, D., & Richardson, A.D. (2016). Using phenocams to monitor our changing Earth: toward a global phenocam network. *Frontiers in Ecology and the Environment, 14*, 84-93

Burkart, A., Hecht, V.L., Kraska, T., & Rascher, U. (2017). Phenological analysis of unmanned aerial vehicle based time series of barley imagery with high temporal resolution. *Precision Agriculture*

Campoy, J.A., Ruiz, D., & Egea, J. (2011). Dormancy in temperate fruit trees in a global warming context: A review. *Scientia Horticulturae, 130*, 357-372

Chuine, I., Kramer, K., & Hänninen, H. (2013). Plant Development Models. In M.D. Schwartz (Ed.), *Phenology: An Integrative Environmental Science* (pp. 275-288). Dordrecht: Springer Netherlands

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 92*, 79-97

D'Odorico, P., Gonsamo, A., Gough, C.M., Bohrer, G., Morison, J., Wilkinson, M., Hanson, P.J., Gianelle, D., Fuentes, J.D., & Buchmann, N. (2015). The match and mismatch between photosynthesis and land surface phenology of deciduous forests. *Agricultural and Forest Meteorology, 214–215*, 25-38

Dandois, J.P., & Ellis, E.C. (2013). High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment, 136*, 259-276

Delbart, N., Beaubien, E., Kergoat, L., & Le Toan, T. (2015). Comparing land surface phenology with leafing and flowering observations from the PlantWatch citizen network. *Remote Sensing of Environment, 160*, 273-280

Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., & Picard, G. (2005). Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sensing of Environment, 97*, 26-38

Delpierre, N., Guillemot, J., Dufrene, E., Cecchini, S., & Nicolas, M. (2017). Tree phenological ranks repeat from year to year and correlate with growth in temperate deciduous forests. *Agricultural and Forest Meteorology, 234*, 1-10

Donnelly, A., Yu, R., Caffarra, A., Hanes, J., Liang, L., Desai, A.R., Liu, L., & Schwartz, M.D. (2017). Interspecific and interannual variation in the duration of spring phenophases in a northern mixed forest. *Agricultural and Forest Meteorology, 243*, 55-67

Duncanson, L.I., Cook, B.D., Hurtt, G.C., & Dubayah, R.O. (2014). An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sensing of Environment, 154*, 378-386

Eastman, J.R., Sangermano, F., Machado, E.A., Rogan, J., & Anyamba, A. (2013). Global Trends in Seasonality of Normalized Difference Vegetation Index (NDVI), 1982–2011. *Remote Sensing, 5*, 4799-4818

Edson, C., & Wing, M.G. (2011). Airborne Light Detection and Ranging (LiDAR) for Individual Tree Stem Location, Height, and Biomass Measurements. *Remote Sensing, 3*, 2494

Elmore, A.J., Guinn, S.M., Minsley, B.J., & Richardson, A.D. (2012). Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. *Global Change Biology, 18*, 656-674

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Forkel, M., Wingate, L., Tomelleri, E., Morra di Cella, U., & Richardson, A.D. (2016). Phenopix: A R package for image-based vegetation phenology. *Agricultural and Forest Meteorology, 220*, 141-150

Fisher, J.I., & Mustard, J.F. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment, 109*, 261-273

Fisher, J.I., Mustard, J.F., & Vadeboncoeur, M.A. (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment, 100*, 265-279

Garrity, S.R., Bohrer, G., Maurer, K.D., Mueller, K.L., Vogel, C.S., & Curtis, P.S. (2011). A comparison of multiple phenology data sources for estimating seasonal transitions in deciduous forest carbon exchange. *Agricultural and Forest Meteorology, 151*, 1741-1752

Hernandez, J.G., Gonzalez-Ferreiro, E., Sarmento, A., Silva, J., Nunes, A., Correia, A.C., Fontes, L., Tomé, M., & Diaz-Varela, R. (2016). Using high resolution UAV imagery to estimate tree variables in Pinus pinea plantation in Portugal. *2016, 25*

Higgins, S.I., Delgado-Cartay, M.D., February, E.C., & Combrink, H.J. (2011). Is there a temporal niche separation in the leaf phenology of savanna trees and grasses? *Journal of Biogeography, 38*, 2165-2175

Hill, R.A., Wilson, A.K., George, M., & Hinsley, S.A. (2010). Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science, 13*, 86-99

Hmimina, G., Dufrêne, E., Pontailler, J.Y., Delpierre, N., Aubinet, M., Caquet, B., de Grandcourt, A., Burban, B., Flechard, C., & Granier, A. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment, 132*, 145-158

Hufkens, K., Friedl, M., Sonnentag, O., Braswell, B.H., Milliman, T., & Richardson, A.D. (2012). Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment, 117*, 307-321

Hunt, E.R., Jr., Hively, W.D., Fujikawa, S.J., Linden, D.S., Daughtry, C.S.T., & McCarty, G.W. (2010). Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. *Remote Sensing, 2*, 290-305

Hunter, A.F., & Lechowicz, M.J. (1992). Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology*, 597-604

Ibáñez, I., Primack, R.B., Miller-Rushing, A.J., Ellwood, E., Higuchi, H., Lee, S.D., Kobori, H., & Silander, J.A. (2010). Forecasting phenology under global warming. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365*, 3247-3260

James, M.R., Robson, S., d'Oleire-Oltmanns, S., & Niethammer, U. (2017). Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment. *Geomorphology, 280*, 51-66

Jönsson, P., Cai, Z., Melaas, E., Friedl, M., & Eklundh, L. (2018). A Method for Robust Estimation of Vegetation Seasonality from Landsat and Sentinel-2 Time Series Data. *Remote Sensing, 10*, 635

Ke, Y., & Quackenbush, L.J. (2011). A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *International Journal of Remote Sensing, 32*, 3625-3647

Keenan, T.F., Darby, B., Felts, E., Sonnentag, O., Friedl, M.A., Hufkens, K., O'Keefe, J., Klosterman, S., Munger, J.W., Toomey, M., & Richardson, A.D. (2014). Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. *Ecological Applications, 24*, 1478-1489

Klosterman, S., Melaas, E., Wang, J., Martinez, A., Frederick, S., O'Keefe, J., Orwig, D.A., Wang, Z., Sun, Q., Schaaf, C., Friedl, M., & Richardson, A.D. (2018). Fine-scale perspectives on landscape phenology from unmanned aerial vehicle (UAV) photography. *Agricultural and Forest Meteorology, 248*, 397-407

Klosterman, S., & Richardson, A. (2017). Observing Spring and Fall Phenology in a Deciduous Forest with Aerial Drone Imagery. *Sensors, 17*, 2852

Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M.A., & Richardson, A.D. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences, 11*, 4305-4320

Lechowicz, M.J. (1984). Why Do Temperate Deciduous Trees Leaf Out at Different Times? Adaptation and Ecology of Forest Communities. *The American Naturalist, 124*, 821-842

Lhermitte, S., Verbesselt, J., Verstraeten, W.W., & Coppin, P. (2011). A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing of Environment, 115*, 3129-3152

Li, D., Guo, H.D., Wang, C., Li, W., Chen, H.Y., & Zuo, Z.L. (2016). Individual Tree Delineation in Windbreaks Using Airborne-Laser-Scanning Data and Unmanned Aerial Vehicle Stereo Images. *IEEE Geoscience and Remote Sensing Letters, 13*, 1330-1334

Liang, L., & Schwartz, M.D. (2009). Landscape phenology: an integrative approach to seasonal vegetation dynamics. *Landscape Ecology, 24*, 465-472
Liang, L., Schwartz, M.D., & Fei, S. (2011). Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. *Remote Sensing of Environment, 115*, 143-157
Lim, Y.S., La, P.H., Park, J.S., Lee, M.H., Pyeon, M.W., & Kim, J.-I. (2015). Calculation of Tree Height and Canopy Crown from Drone Images Using Segmentation. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 33*, 605-614
Lin, C., Chen, S.-Y., Chen, C.-C., & Tai, C.-H. (2018). Detecting newly grown tree leaves from unmanned-aerial-vehicle images using hyperspectral target detection techniques. *ISPRS Journal of Photogrammetry and Remote Sensing, 142*, 174-189
Lisein, J., Michez, A., Claessens, H., & Lejeune, P. (2015). Discrimination of Deciduous Tree Species from Time Series of Unmanned Aerial System Imagery. *PLoS ONE, 10*
Liu, L., Liang, L., Schwartz, M.D., Donnelly, A., Wang, Z., Schaaf, C.B., & Liu, L. (2015). Evaluating the potential of MODIS satellite data to track temporal dynamics of autumn phenology in a temperate mixed forest. *Remote Sensing of Environment, 160*, 156-165
Liu, Y., Hill, M.J., Zhang, X., Wang, Z., Richardson, A.D., Hufkens, K., Filippa, G., Baldocchi, D.D., Ma, S., Verfaillie, J., & Schaaf, C.B. (2017). Using data from Landsat, MODIS, VIIRS and PhenoCams to monitor the phenology of California oak/grass savanna and open grassland across spatial scales. *Agricultural and Forest Meteorology, 237–238*, 311-325
Mei, C., & Durrieu, S. (2004). Tree crown delineation from digital elevation models and high resolution imagery. *Proc. IAPRS, 36*, 218-223
Melaas, E.K., Friedl, M.A., & Zhu, Z. (2013). Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM plus data. *Remote Sensing of Environment, 132*, 176-185
Melaas, E.K., Sulla-Menashe, D., Gray, J.M., Black, T.A., Morin, T.H., Richardson, A.D., & Friedl, M.A. (2016). Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. *Remote Sensing of Environment, 186*, 452-464
Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-KÜBler, K., Bissolli, P., Braslavská, O.g., & Briede, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology, 12*, 1969-1976
Miao, L., Luan, Y., Luo, X., Liu, Q., Moore, J.C., Nath, R., He, B., Zhu, F., & Cui, X. (2013). Analysis of the Phenology in the Mongolian Plateau by Inter-Comparison of Global Vegetation Datasets. *Remote Sensing, 5*, 5193-5208
Michez, A., Piégay, H., Lisein, J., Claessens, H., & Lejeune, P. (2016). Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. *Environmental Monitoring and Assessment, 188*, 1-19
Mizunuma, T., Wilkinson, M., L Eaton, E., Mencuccini, M., Il Morison, J., & Grace, J. (2013). The relationship between carbon dioxide uptake and canopy colour from two camera systems in a deciduous forest in southern England. *Functional Ecology, 27*, 196-207
Moore, C.E., Brown, T., Keenan, T.F., Duursma, R.A., van Dijk, A., Beringer, J., Culvenor, D., Evans, B., Huete, A., Hutley, L.B., Maier, S., Restrepo-Coupe, N., Sonnentag, O., Specht, A., Taylor, J.R., van Gorsel, E., & Liddell, M.J. (2016). Reviews and syntheses: Australian vegetation phenology: new insights from satellite remote sensing and digital repeat photography. *Biogeosciences, 13*, 5085-5102
Morin, X., Roy, J., Sonie, L., & Chuine, I. (2010). Changes in leaf phenology of three European oak species in response to experimental climate change. *New Phytologist, 186*, 900-910

Morisette, J.T., Richardson, A.D., Knapp, A.K., Fisher, J.I., Graham, E.A., Abatzoglou, J., Wilson, B.E., Breshears, D.D., Henebry, G.M., Hanes, J.M., & Liang, L. (2009). Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Frontiers in Ecology and the Environment, 7*, 253-260

Morris, D.E., Boyd, D.S., Crowe, J.A., Johnson, C.S., & Smith, K.L. (2013). Exploring the Potential for Automatic Extraction of Vegetation Phenological Metrics from Traffic Webcams. *Remote Sensing, 5*, 2200-2218

Nasahara, K., & Nagai, S. (2015). Review: Development of an in situ observation network for terrestrial ecological remote sensing: the Phenological Eyes Network (PEN). *Ecological Research, 30*, 211-223

Newcastle University (2015). Cockle Park Weather Station monthly records. In

Norby, R.J., Hartz-Rubin, J.S., & Verbrugge, M.J. (2003). Phenological responses in maple to experimental atmospheric warming and CO2 enrichment. *Global Change Biology, 9*, 1792-1801

Panagiotidis, D., Abdollahnejad, A., Surový, P., & Chiteculo, V. (2016). Determining tree height and crown diameter from high-resolution UAV imagery. *International Journal of Remote Sensing*, 1-19

Peng, D.L., Zhang, X.Y., Zhang, B., Liu, L.Y., Liu, X.J., Huete, A.R., Huang, W.J., Wang, S.Y., Luo, S.Z., Zhang, X., & Zhang, H.L. (2017). Scaling effects on spring phenology detections from MODIS data at multiple spatial resolutions over the contiguous United States. *ISPRS Journal of Photogrammetry and Remote Sensing, 132*, 185-198

Polgar, C.A., & Primack, R.B. (2011). Leaf-out phenology of temperate woody plants: from trees to ecosystems. *New Phytologist, 191*, 926-941

Polgar, C.A., & Primack, R.B. (2013). Leaf out phenology in temperate forests. *Biodiversity Science, 21*, 111-116

Richardson, A., & O'Keefe, J. (2009). Phenological Differences Between Understory and Overstory. In A. Noormets (Ed.), *Phenology of Ecosystem Processes* (pp. 87-117): Springer New York

Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., & Ollinger, S.V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications, 19*, 1417-1428

Rodriguez-Galiano, V.F., Dash, J., & Atkinson, P.M. (2015). Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. *Geophysical Research Letters, 42*, 2253-2260

Ryan, C.M., Williams, M., Hill, T.C., Grace, J., & Woodhouse, I.H. (2012). Assessing the Phenology of Southern Tropical Africa: A Comparison of Hemispherical Photography, Scatterometry, and Optical/NIR Remote Sensing

Ryu, Y., Verfaillie, J., Macfarlane, C., Kobayashi, H., Sonnentag, O., Vargas, R., Ma, S., & Baldocchi, D.D. (2012). Continuous observation of tree leaf area index at ecosystem scale using upward-pointing digital cameras. *Remote Sensing of Environment, 126*, 116-125

Samiappan, S., Turnage, G., Hathcock, L.A., & Moorhead, R. (2017). Mapping of invasive phragmites (common reed) in Gulf of Mexico coastal wetlands using multispectral imagery and small unmanned aerial systems. *International Journal of Remote Sensing, 38*, 2861-2882

Schaber, D.S.-W.J. (2002). Phenology in Germany in the 20th century: methods, analyses and models. In, *Department of Geoecology* (p. 146): UNIVERSITY of POTSDAM, GERMANY

Schaber, J., & Badeck, F.-W. (2003). Physiology-based phenology models for forest tree species in Germany. *International Journal of Biometeorology, 47*, 193-201

Schwartz, M.D., Hanes, J.M., & Liang, L. (2013). Comparing carbon flux and high-resolution spring phenological measurements in a northern mixed forest. *Agricultural and Forest Meteorology, 169*, 136-147

Smith, G.M., & Milton, E.J. (1999). The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing, 20*, 2653-2662

Smith, R.J. (2009). Use and Misuse of the Reduced Major Axis for Line-Fitting. *American Journal of Physical Anthropology, 140*, 476-486

Snavely, N., Seitz, S.M., & Szeliski, R. (2008). Modeling the world from Internet photo collections. *International Journal of Computer Vision, 80*, 189-210

Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A.M., Friedl, M., Braswell, B.H., Milliman, T., O'Keefe, J., & Richardson, A.D. (2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology, 152*, 159-177

Sparks, T.H. (2014). Local-scale adaptation to climate change: the village flower festival. *Climate Research, 60*, 87-89

St-Onge, B., Audet, F.A., & Begin, J. (2015). Characterizing the Height Structure and Composition of a Boreal Forest Using an Individual Tree Crown Approach Applied to Photogrammetric Point Clouds. *Forests, 6*, 3899-3922

Storey, J., Choate, M., & Lee, K. (2014). Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance. *Remote Sensing, 6*, 11127-11152

Tang, J.W., Korner, C., Muraoka, H., Piao, S.L., Shen, M.G., Thackeray, S.J., & Yang, X. (2016). Emerging opportunities and challenges in phenology: a review. *Ecosphere, 7*

Thackeray, S.J., Sparks, T.H., Frederiksen, M., Burthe, S., Bacon, P.J., Bell, J.R., Botham, M.S., Brereton, T.M., Bright, P.W., Carvalho, L., Clutton-Brock, T.I.M., Dawson, A., Edwards, M., Elliott, J.M., Harrington, R., Johns, D., Jones, I.D., Jones, J.T., Leech, D.I., Roy, D.B., Scott, W.A., Smith, M., Smithers, R.J., Winfield, I.J., & Wanless, S. (2010). Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology, 16*, 3304-3313

Thiel, C., & Schmullius, C. (2016). DERIVATION OF FOREST PARAMETERS FROM STEREOGRAPHIC UAV DATA–AComparison WITH AIRBORNE LIDAR DATA. In L. Ouwehand (Ed.), *Living Planet Symposium* (p. 189). Prague, Czech Republic: ESA-SP

Thompson, R., & Clark, R.M. (2008). Is spring starting earlier? *The Holocene, 18*, 95-104

Torresan, C., Andrea Berton, & Federico Carotenuto, S.F.D.G., Beniamino Gioli, Alessandro Matese, Franco Miglietta, Carolina Vagnoli, Alessandro Zaldei & Luke Wallace (2017). Forestry applications of UAVs in Europe: a review. *International Journal of Remote Sensing, 38*, 2427-2447

Tsingas, V. (1992). *Automatisierung der Punktubertragung in der Aerotriangulation durch mehrfache digitale Bildzuordnung*. Munchen

USGS (2016a). Landsat 4-7 climate data record (CDR) surface reflectance. In, *Product guide* (p. 42): U.S. Geological Survey

USGS (2016b). Provisional landsat 8 surface reflectance code (LaSRC) product. In, *Product guide* (p. 27): U.S. Geological Survey

Verhoeven, G.J. (2012). Near-Infrared Aerial Crop Mark Archaeology: From its Historical Use to Current Digital Implementations. *Journal of Archaeological Method and Theory, 19*, 132-160

Vitasse, Y., Delzon, S., Dufrêne, E., Pontailler, J.-Y., Louvet, J.-M., Kremer, A., & Michalet, R. (2009). Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology, 149*, 735-744

Wang, Q.M., Blackburn, G.A., Onojeghuo, A.O., Dash, J., Zhou, L.Q., Zhang, Y.H., & Atkinson, P.M. (2017). Fusion of Landsat 8 OLI and Sentinel-2 MSI Data. *IEEE Transactions on Geoscience and Remote Sensing, 55*, 3885-3899

White, K., Pontius, J., & Schaberg, P. (2014). Remote sensing of spring phenology in northeastern forests: A comparison of methods, field metrics and sources of uncertainty. *Remote Sensing of Environment, 148*, 97-107

White, M.A., de Beurs, K.M., Didan, K., Inouye, D.W., Richardson, A.D., Jensen, O.P., O'Keefe, J., Zhang, G., Nemani, R.R., van Leeuwen, W.J.D., Brown, J.F., de Wit, A., Schaepman, M., Lin, X., Dettinger, M., Bailey, A.S., Kimball, J., Schwartz, M.D., Baldocchi, D.D., Lee, J.T., & Lauenroth, W.K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology, 15*, 2335-2359

Wilkinson, M., Eaton, E.L., Broadmeadow, M.S.J., & Morison, J.I.L. (2012). Inter-annual variation of carbon uptake by a plantation oak woodland in south-eastern England. *Biogeosciences, 9*, 5373-5389

Woodget, A.S., Austrums, R., Maddock, I.P., & Habit, E. (2017). Drones and digital photogrammetry: from classifications to continuums for monitoring river habitat and hydromorphology. *Wiley Interdisciplinary Reviews: Water*, e1222-n/a

Xin, Q.C., Olofsson, P., Zhu, Z., Tan, B., & Woodcock, C.E. (2013). Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. *Remote Sensing of Environment, 135*, 234-247

Zaki, N.A.M., Latif, Z.A., Zainal, M.Z., & Zainuddin, K. (2015). Individual tree crown (ITC) delineation using watershed transformation algorithm for tropical lowland dipterocarp. In, *2015 International Conference on Space Science and Communication (IconSpace)* (pp. 237-242)

Zhang, X. (2015). Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data. *Remote Sensing of Environment, 156*, 457-472

Zhang, X., Friedl, M., & Schaaf, C. (2009). Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. *International Journal of Remote Sensing, 30*, 2061-2074

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment, 84*, 471-475

Zhang, X.Y., Wang, J.M., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y.Y., Jayavelu, S., Gray, J., Liu, L.L., Yan, D., & Henebry, G.M. (2017). Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sensing of Environment, 190*, 318-330

Zhao, J., Zhang, Y., Tan, Z., Song, Q., Liang, N., Yu, L., & Zhao, J. (2012). Using digital cameras for comparative phenological monitoring in an evergreen broad-leaved forest and a seasonal rain forest. *Ecological Informatics, 10*, 65-72

8. List of Figure Captions

Figure 1. Frequency of data collection/data availability over the study area. The total number of observations are given by 'n'. Vertical lines represent the date when the first tree was observed to reach SOS and the date when the last tree reached EOS. The ground data collection was intensified from ~DOY 90 to ~DOY 130 in order to better monitor SOS.

Figure 2. Effect of the 80th percentile method on the smoothed UAV GCC_{DN} time series data (black dots) and consequent phenometric estimations in the Sycamore plot. The data are fitted by the greendown model (solid line), with phenometrics (SOS, MOS and EOS) marked by the vertical dashed lines. The model failed to estimate phenometrics in a). The averaged SOS from visual assessments is also shown for comparison purposes.

Figure 3. Visually assessed dates compared against dates estimated from the UAV remote sensing derived products, at an individual-tree level. UAV phenometrics are calculated based on the least uncertain sigmoid model per plot (Table S1; Sycamore, Sitka s., Norway s. and Mix = Greendown model; Larch and Oak = Simple model). **p<0.001, *p<0.05; N is sample size. Bias refers to the average difference between UAV and visual observations. Dashed lines represent the 1:1 line and solid lines are RMA regression models. Statistics are given in Table 2.

Figure 4. Plot level comparison between Landsat (a-l), UAV (m-r) and understorey ground photography (s-x). The time series of data (black dots) are fitted by the "best model" per plot (Table S1 to Table S4), with phenometrics marked by the vertical dashed lines (preceded by 'Fit_' in the legend). The averaged SOS from visual assessments and the UAV SOS are also shown for comparison purposes. Landsat MOS and EOS metrics are shown in this figure, but are not considered in the comparisons with the UAV phenology (these metrics are likely not meaningful due to the large Landsat data gap from DOY 113-161).

Figure 5. a) Location of the study area and individual tree-level predictions of SOS using UAV GCC_{DN} (80th percentile; fitted by the greendown model), and b) pixel-level predictions of SOS using Landsat NDVI (fitted by the simple model). Background orthomosaic derived from UAV images (visible camera) acquired on 21/04/2015 (DOY 111).

Figure 6. Boxplots of SOS dates of individual trees (n = 4354), as mapped in Figure 5a. For each boxplot, the central mark represents the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers (<2 times the standard deviation). Broadleaf encompasses the Sycamore, Oak and Mix plots.

Figure 7. Comparison between UAV SOS and Landsat SOS. Landsat LSP (simple model) compared with: a-b) UAV-derived (greendown model) averaged tree-level phenology (within each Landsat pixel area); and c-f) with UAV LSP (greendown model) considering the entire Landsat pixel area. Pure deciduous covers are shown separately (top row). Dashed lines represent the 1:1 line and solid lines are RMA regression models (**p<0.001; *p<0.05). Statistics are given in Table 3.

Figure 8. Comparison between UAV- and Landsat-based VI time series, at a Landsat pixel scale, considering the entire woodland (a-d) and only the deciduous cover (e-h). UAV VIs are based on the mean of the whole polygon (Landsat grid). Dashed lines represent the 1:1 line and solid lines are RMA regression models. Statistics are given in Table 4.